AUTOMATIC VIDEO SIMILARITY MEASURE USING VECTORIZED ATTRIBUTED GRAPH MATCHING

Miss Suprachaya Veeraprasit

A Dissertation Submitted in Partial Fulfillment of the Requirements
for the Degree of Doctor of Philosophy Program in Computer Science and
Information Technology
Department of Mathematics and Computer Science
Faculty of Science
Chulalongkorn University
Academic Year 2016

การวัดความคล้ายของวิดีทัศน์แบบอัตโนมัติโดยใช้การจับคู่กราฟลักษณะประจำแบบเวกเตอร์

นางสาวสุปรัชญา วีระประสิทธิ์

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรดุษฎีบัณฑิต
สาขาวิชาวิทยาการคอมพิวเตอร์และเทคโนโลยีสารสนเทศ ภาควิชาคณิตศาสตร์และวิทยาการคอมพิวเตอร์
คณะวิทยาศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย
ปีการศึกษา 2559

Thesis Title	AUTOMATIC VIDEO SIMILARITY MEASURE USING VECTORIZED ATTRIBUTED GRAPH MATCHING
By	Miss Suprachaya Veeraprasit
Field of Study	Computer Science and Information Technology
Thesis Advisor	Assistant Professor Suphakant Phimoltares, Ph.D.
Thesis Co-Advisor	Professor Chidchanok Lursinsap, Ph.D.

Accepted by the Faculty of Science, Chulalongkorn University in Partial Fulfillment of the Requirements for the Doctoral Degree

.....Dean of the Faculty of Science
(Associate Professor Polkit Sangvanich, Ph.D.)

THESIS COMMITTEE

.....Chairman
(Assistant Professor Annupan Rodtook, Ph.D.)
.....Thesis Advisor
(Assistant Professor Suphakant Phimoltares, Ph.D.)
.....Thesis Co-Advisor
(Professor Chidchanok Lursinsap, Ph.D.)
.....Examiner
(Assistant Professor Saranya Maneeroj, Ph.D.)
.....Examiner
(Associate Professor Chatchawit Aporntewan, Ph.D.)
.....External Examiner
(Assistant Professor Jakkarin Suksawatchon, Ph.D.)

สุปรัชญา วีระประสิทธิ์ : การวัดความคล้ายของวิดีทัศน์แบบอัตโนมัติโดยใช้การจับคู่กราฟลักษณะประจำแบบเวกเตอร์ (AUTOMATIC VIDEO SIMILARITY MEASURE USING VECTORIZED ATTRIBUTED GRAPH MATCHING) อ.ที่ปรึกษาวิทยานิพนธ์หลัก: ผศ. ดร. ศุภกานต์ พิมลธเรศ, อ.ที่ปรึกษาวิทยานิพนธ์ร่วม: ศ. ดร. ชิดชนก เหลือสินทรัพย์, 99 หน้า.

การวัดความคล้ายของฉากในวิดีทัศน์ได้ถูกแนะนำให้อยู่ในกลุ่มของการวิเคราะห์ตำแหน่งของวัตถุในฉากหลังที่ได้รับการสกัดออกมาจากฉากของวิดีทัศน์นั้น ได้ถูกแปลงรูปแบบให้อยู่ในรูปแบบของฟังก์ชันตรรกะเชิงพื้นที่ ฟังก์ชันเชิงตรรกะด้านพื้นที่ที่ได้ใช้ในการแสดงแทนตำแหน่งของวัตถุบนฉากได้ถูกถอดรหัสให้อยู่ในรูปแบบของแบบจำลองทรรศนะด้านบนที่ถูกนำเสนอบนระนาบแบบตารางหน่วยในภายหลัง จากระนาบแบบตารางหน่วยที่ถูกจำลอง สารสนเทศข้อมูลแบบจำลองทรรศนะด้านบนของฉากวิดีทัศน์ได้ถูกเก็บในรูปแบบดิจิทัลที่สามารถนำมาใช้ได้อย่างหลากหลายและง่าย นอกจากนี้ข้อมูลเหล่านั้นที่อยู่ในระนาบแบบตารางหน่วยยังเป็นส่วนสำคัญในกราฟแสดงความหมายเชิงพื้นที่การแปลงจากตารางหน่วยไปเป็นกราฟนั้นได้ถูกอธิบายไว้ในวิทยานิพนธ์ฉบับนี้ กราฟแสดงความหมายเชิงพื้นที่ได้ถูกใช้เพื่อระบุแบบอย่างของตำแหน่งของวัตถุในฉาก เนื่องจากลักษณะเฉพาะของกราพแสดงความหมายเชิงพื้นที่ การเปรียบเทียบระหว่างฉากวิดีทัศน์จึงถูกทำให้ง่ายขึ้นการแปลงไปเป็นแบบจำลองทรรศนะด้านบนนั้นทนทานต่อทัศนมิติ ถึงแม้ว่าทรรศนะจะถูกเปลี่ยนแปลงตามทัศนมิติ แต่กราฟแสดงความหมายเชิงพื้นที่สามารถบ่งชี้ความคล้ายผ่านแบบจำลองทรรศนะด้านบนได้ ความทนทานของขั้นตอนวิธีนี้ได้ถูกแสดงโดยตัวอย่างของการทดลองบางส่วน

ภาควิชา	คณิตศาสตร์และวิทยาการคอมพิวเตอร์	ลายมือชื่อนิสิต .....
สาขาวิชา	วิทยาการคอมพิวเตอร์และเทคโนโลยีสารสนเทศ	ลายมือชื่อ อ.ที่ปรึกษาหลัก .....
ปีการศึกษา	2559	ลายมือชื่อ อ.ที่ปรึกษาร่วม .....

5473104423 : MAJOR COMPUTER SCIENCE AND INFORMATION TECHNOLOGY
KEYWORDS: SEMANTIC SPATIAL GRAPH / TOP-VIEW MODEL / VIDEO SCENE SIMILARITY

SUPRACHAYA VEERAPRASIT: AUTOMATIC VIDEO SIMILARITY MEASURE USING VECTORIZED ATTRIBUTED GRAPH MATCHING. ADVISOR: ASST. PROF. SUPHAKANT PHIMOLTARES, Ph.D., CO-ADVISOR: PROF. CHIDCHANOK LURSINSAP, Ph.D., 99 pp.

The video scene similarity is introduced in term of the objects' positions analysis. The background objects' positions which are extracted from the video scene are transformed into the spatial logical functions. The spatial logical function which is used to represent the objects' positions on the scene are decoded into top-view model which is projected on the grid unit plane later. From the simulated grid unit plane, the video scene top-view model information are kept in digital form which can be used variously and easily. Besides those data in grid unit plane are the features in the semantic spatial graph. The transformation from unit plane into graph is described in this dissertation. The semantic spatial graph is used to identify the pattern of the objects' position in a scene. Due to the characteristic of semantic spatial graph, the comparing between video scenes are simplified. The transformation to top-view model is robust to perspective. Eventhough the view are changed perspectively, the semantic spatial graph can determined the similarity through top-view model. The robustness of this algorithm are showed by the some examples of the experiments.

Department:	Mathematics and Computer Science	Student's Signature .....
		Advisor's Signature .....
Field of Study:	Computer Science and Information Technology	Co-Advisor's Signature .....
Academic Year:	2016	

ACKNOWLEDGEMENTS

I feel greatful for my advisor and co-advisor who had paid much effort to complete this research.

Thank you very much for who had paid effort and help to complete this research.

CONTENTS

Page

LIST OF TABLES

LIST OF FIGURES

1. INTRODUCTION

Concluding whether two scenes consisting of the same set of objects with arrangement and different viewing angles are from the same location or not is not simple. Given a set of objects, different configurations of arrangement patterns can give different interpretations of the scenes. For example, arranging five chairs around a round table has different meaning from arranging five chairs in a row with the round table behind the row of chairs. This problem has wide applications in various fields. Some of the applications are crime scene identification and exact scene retrieval. The challenging issues of this problem are how to determine the relatively identical arrangement of objects in two scenes and where physical locations of objects in the scenes are. In general, this problem differs from the problem of image retrieval and video retrieval where a set of similar images or videos are grouped together by some measures. Those retrieved images or videos are not exactly from the same image or video with different view angles. They are usually from different places or locations having similar sets of objects.

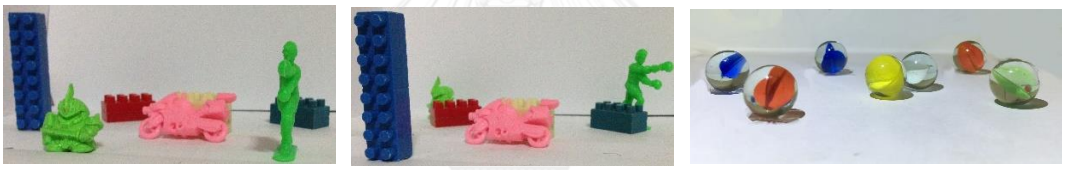


Figure 1. Example of three scenes with different arrangements.

Video classification is one of popular research categories in computer vision as shown in a survey of video classification [1]. Illumination and objects in videos are typically used to classified video into desired categories. Particularly, objects in video scenes are mainly focused to indicate the categories and similarity. Normally in human vision, people recognize scenes by the objects scattered in the scenes. However, a computer senses only digital information and interprets the information as objects by computations. There are various related researches about the image segmentation, object distinction, semantic understanding for computer and robots that are significant to handle such tasks. The content-based video retrieval (CBVR) problem also used the objects in a video for information retrieving system. Objects are recognized based on the interested domain. Jones and Shao [2] focused in human-action domain for video databases retrieving by searching the descriptions. Besides the contents in the videos, the caption of video is required. For semantics, the video annotation is related problem. Ding et al. [3] proposed the video annotation framework using bag of words. Likewise, there was a framework using bag of events proposed by Zhang et al. [4] Moreover, the relationship between the detected objects can assist the analysis in computer vision. Advancing in the video retrieval, the relationship between objects in videos is a challenging retrieval method and feature extraction. There are some situations using low level feature such as histogram to differentiate two scenes. But this approach is not

efficient enough. For example, picture on the left and middle scenes of Figure 1. have the same set of objects but different configurations of object arrangement. These two scenes cannot be interpreted as the same scene because of the arrangement of objects. However, if the methods of image retrieval are applied, then these two scenes are considered as similar scenes. Arrangement implies the semantic distinction between scenes. Moreover, the attributes of objects, such as color, shape, texture and semantic, are important in considering the semantic distinction. The right scene of Figure 1. has the same arrangement and number of objects to left scene. If the objects' attributes are not used in comparing, those two scenes can be considered as a same scene with the same arrangement. Thus, both attributes and arrangement of objects are significant in comparing semantic between two scenes.

Recognized objects and relations of object arrangement can be represented by graph. A lot of researches used graphs for pattern matching such as Conte et al. [5] The significant part for graph analysis is the graph matching. Pawar and Zaveri [6] proposed the pattern matching for shape recognition by using error-correcting or graph edit distance. Even the web-pages and links can be represented by a directed graph and transformed it into a matrix for enhancing the search engine [7]. To find the good and interesting pages among the web-pages when query, the authorities score is applied to matching for evaluate similarity score. Such large graphs like web-pages' link are easier to be measured by scoring, edit distance or subgraph matching, as mentioned by Zager and Verghese [8]. Until now, there are a number of graph matching algorithms, therefore representing the video scenes or images by graph is commonly used. Chevalier et al. [9], considered the video scenes as region adjacency graph (RAG) depending on the containing objects. Each region or vertex contains color and shape as the attributes for calculating the similarity between RAGs. Besides, the objects are detected. Using graph to represent the objects and relationship between them might serve further purposes in semantic ways. As mentioned, to support high level features in query of the video scenes, the scene similarity with the graph that represented objects and their relationships is necessary.

In this work, the similarity is analyzed by the objects' positions. The relations between the positions of objects serve as features for scene similarity analysis. Naturally, humans recognize scenes by their memories of the objects that they have seen. Our algorithm is adapted from this idea. When human saw the scene, they perceived the objects' positions in top-view model. After they saw the same scene in other perspective of view, they still recognize them as the same scene which they already saw. The size of object also plays the significant role in recognizing. People recognize the same place by observing the significant objects and their positions even they look from different perspective. On the hypothesis, even the perspective is changed, the objects' position relationships are not changed. The similarity can be determined by the objects' position. Using the algorithm in this dissertation, the top-

view of scenes are extracted. The structure of objects in top-view model can be presented in form of objects' relation graphs. The objects' relation graphs are used to determine the similarity between scenes in the dataset. The graph that represent the objects' relation is called the Vectorized Attributed Graph (VAG). For this reason, two scenes can be determined that they are from the same place with the same structure of objects around the place. Even some objects are missing or misplace, Vectorized Attributed Graph can be useful to determine the similarity.

Further in indoor scene, there are multi-surfaces with groups of objects on them. Thus detecting the surfaces conduces to detecting objects that are on the surfaces. There are quite some methods to detect surfaces or planes in the images. Instead of detecting objects first, the surfaces or planes are detected at the beginning. If objects are on surfaces, the objects that are not on the detected surfaces might cause false alarms. After the surfaces are detected, a model to represent the objects with surfaces on another surface such as a table on a floor is needed. To organize multi-level surfaces, hierarchical graph is considered. The planes are encoded into top-view models. Later for more than one plane, a top-view model acts as a level of hierarchical graph. In conclusion, the Vectorized Attributed Graph (VAG) is the hierarchical graph that has the vectorized attributed nodes. The objectives of the dissertation along with the problem formulation are as follow.

1.1. Objectives

The main objectives of this dissertation are as follow:

1. To assign the attributes from each detected objects into a node of Vectorized Attributed Graph.
2. To develop a method to measure scene similarity using Vectorized Attributed Graph.

1.2. Problem Formulation

Given the two scenes s_i and s_j with the same number of objects, determine whether s_i and s_j are similar or not. This problem can be decomposed into following sub-problems.

1. How to capture or represent the video scene in semantic way using Vectorized Attributed Graph (VAG)?
2. How to determine the similarity between two scenes that are represented by Vectorized Attributed Graph (VAG)?

From the above, this study is concentrated in constructing the Vectorized Attributed Graph (VAG). The interesting inputs are the detected objects on the video input. To compare the similarity between video sequences in semantic way, the layout of appeared objects are significant. In this experiment, the objects are obtained from the

existing object detection method. Those objects are labeled with the semantic. Those methods are described in the theoretical background section.

1.3. Scope of the work

In this research, the experiments have some constraints as follows:

1. Video without moving objects.
2. Frames of interest are main components to be analyzed.
3. Offline processing.
4. Indoor scene.

In this dissertation, related researches are shown in next chapter. Next, theoretical background used in the experiments is described in chapter 3. Subsequently, the proposed methods are shown in chapter 4 along with the experimental results. Later the videos are extracted into frames. Each video produced many frames. The video frame comparison method is introduced based on the fact that videos are image sequences. Finally, the last chapter presents conclusion of this research.

2. LITERATURE REVIEW

In computer vision, to interpret the image or video for analyzing into computer term is the main problem. For video similarity and comparison, there are quite many researches such as [9-11]. Their proposed work is based on segmenting a frame into region by color. Color is a low-level feature which might not serve specific purpose as querying videos by the objects' name. In present, the annotation on the video research is expanded. Su et al. [12] stated a content-based video retrieval method. Their method inspired the idea for further use in future. Yang et al. [13] proposed object recognition method using bag of word for helping the robot. These researches show the possibility that the objects can be distinctive from the video scenes. Further, improvement on the video comparison and retrieval is risen. The high-level features are considered to be useful in the semantic level of comparing or searching of videos. Shearer et al. [14, 15] proposed the video indexing and similarity retrieval using the spatial relationship to represent the objects. The indexing using graph is used in video databases retrieval problem. Arndt et al. [16] proposed image sequence comparison based on the object relation. The relationship between objects that appear in video scene served as high level features in video analysis. From this view point, a graph is useful to express this relationship between objects. The applications involving graph are described in next paragraph.

Graph is a structure variously used to solve many problems. Especially, the problems have the relationship among data themselves or can be interpreted into graphs. There are a lot of mathematical or biological applications using graph as a key structure for analyzing data and finding a solution. The graph can represent data and recognize the necessary information. Also, the graph algorithm is strongly required for serving various purposes. In biological application, a graph represents protein-protein interactions (PPIs) and their networks, as shown in Rohit Singh et al.'s research[17]. This data representation plays the important role to discover and understand the interactions between proteins. From their work, representing vertex as protein and edge as direct physical interaction between proteins yields satisfaction in analysis. The IsoRank algorithm is first proposed in this article. Their objective is to find the size of common graph from mapping between two graphs and the sequence similarity between vertices mapped among a pair of graphs. Likewise, the graph algorithm is useful in other science. Basically a graph contains vertices and edges. For using graph to represent data, the graph analysis method is applied such as the graph comparing. As describing further in this section, a graph presents objects and their relationship in the scene. For analysis between scenes, the graph similarity algorithm is adapted to solve the problem of scene similarity. Next, three categories for the related works are described as follows.

2.1. Object Scene in Computer Vision

In computer vision, works involving object recognition is quite notable. To represent objects in scene for further use, various methods including graph algorithm are applied. There are many applications using object recognition concept. To understand and remember the picture as same as human vision is a problem that have been researched and developed. Lowe et al.'s showed that objects can be recognized and detected distinctively in a scene [18, 19]. Thus, analyzing the scene containing objects is another line of computer vision that have been researched thoroughly. Mekhalfi et al. [20] proposed the comparison among using PCA, SIFT, and bag of words as features, to recognize the objected in scene for blind people. Those algorithms use the feature extraction with the second level-feature library. This library is used for image matching to define the descriptors of each scene. Then extracted features are fed into analysis and clustering algorithm to find the result of object list in a scene. Their purpose is to tell the blind people where they are. Therefore, the objects in scene is the necessity in comparing between videos or images.

To distinguish between scenes, the distinct objects are proved to be useful. As Shearer et al.'s research [14], the distinct objects are rewritten into top view model from each key frame in the video to produce video index. This shows how important the distinct objects in each scene can differentiate a scene from another scene. If objects can be practically recognized, those can be used as the high-level features. These features can be used in semantic ways of recognizing objects.

2.2. Scene Similarity

The purpose is to measure similarity between scenes for both images and videos. This category applied to video analysis involving summarization, annotation, comparison and indexing. The video summarization and indexing can be grouped together. The video comparison is the sub-problem of those summarization and indexing. If the video comparison problems are performed, the others should be done as well. Lastly, the video annotation can be counted as the part of video comparison. Sun et al. [21] proposed the annotation method to specify the vehicle in web videos. In their research, the moving objects are captured along with their audio as extracted features. Gang and Xiaochi[11] proposed the scene recognition which used bag of words to create histogram as codebook for each scene description. Cao and Zhu[10] proposed the video similarity search algorithm. They used color histogram to make the image characteristic code as their features for looking up their database of codes. In video classification, Xu and Li [22] used PCA to analyze the audio-visual features and compare with their spatial-temporal audio-visual feature vectors. It can be concluded that scene similarity is important to solve the other problems.

2.3. Graph Similarity

Zhou and Torre [23] proposed the advantage of the graph matching method in recognizing the objects. In addition, they presented factorized graph matching algorithm with optimization to improve the graph matching performance. One of the problems of graph matching is the density of the graph vertices. High density causes the complexity and time consumption in calculation. Lee et al. [24] proposed video abstractions using graph matching. They similarly proposed their graph similarity measure. Video frame is segmented according to the color regions. Then, those regions form RAG. Their graph similarity measure is calculated from the number of neighbor subgraphs. Then the graph of differences is plotted and the gradual different graph similarity value is found to collect the frame into video abstract.

As previously mentioned, graph is practically used in video analysis. The graph structure represents the scene effectively. Thus it can be said that scene comparison is equivalent to graph comparison. There are quite number of researches for graph matching and graph similarity. The general algorithm for measuring the similarity of graphs is graph isomorphism. Due to the time consumption, the common subgraph isomorphism is more efficient to measure graph similarity. Dahm et al.[25] presented the benefit of topological node features (TNFs). Furthermore, they presented more advanced topological node features which are n-neighborhood feature. The n-neighborhood features utilized the existing TNFs by developing the induced subgraphs formed from all nodes that can be reached within n steps. This work emphasizes the significance of the relationship in neighborhood of objects or nodes.

2.4. Plane Detection

Usually, an object lies on a surface or a plane such as a table front, a desk or a counter. In many indoor scenes, the floor or walls are the most of the area in a picture. By detecting the floor, objects that are on the floor are also detected. Bao et al. [26] proposed the idea which using the geometric contextual reasoning for object recognition. They found the fact that objects must lie on the supportive surfaces. The relationship between objects and planes are formed in the 3D scene layout according to camera pose, focal length and location. They proved that their models have ability to reduce false alarm and false negative object detection rate. As in their assumptions, the objects must lie on a surface or a plane and on upright pose. In their paper, the objects in the image are measured size, camera distance and angle that objects reflect with surface. In their method, they show that single camera images can be used in extracting objects along with multi-planes. The planes in indoor scenes can be detected by using the lines appeared. In Ramalingam's paper [27], they used Manhattan junction ideas, after line segmentations into tree directions. They identified the junction into three

types, such as L, T, W, Y and X. They proved that their method is efficient in using them for spatial recognition of indoor scenes. In Park's work, they proposed a framework that using a single image to recover a 3D cuboidal indoor scene with a semantic segmentation feature. They introduce the discrimination between objects and room faces using labeling energy. The neighboring superpixels based on angles from vanishing points are encoded. The differences of colors are used to describe the differences of surfaces. They also use the relation of vanishing points between two different surfaces in labeling their energies. The methods, they provided proved that objects can be detected using the plane in single image.

2.5. Hierarchical graph

There are possibilities that an indoor scene is extracted with multiple planes or surfaces. Those multi-planes will be represented with a specific model to show their relationship such as a surface on another surface. For example, a table on the floor with a group of objects on it. Normally, a group of objects might be represented by a graph where nodes represent objects and edges are the relationship between objects. Tree also is a kind of graph with multi-level representation. Normally, there are many algorithms to solve the matching problem of this kind of graph. Shokoufandeh proposed the matching of hierarchical features in image [28]. They extracted saliency regions from an image as nodes of saliency map graph. For their saliency map graph, the graph contained both topological and geometric information. The multiscale wavelet transform was applied to image. The hierarchical map chose salient regions at the appropriate scales of resolution. Each region mapped to a node in a directed acyclic graph. The larger region that covers smaller region was represented as one parent node. From their proposed graph, the bijective mapping method was used in their experiments to match saliency map graph which is a generalization of the graph isomorphism problem. In the matching algorithm, they found a maximum cardinality and minimum weight mapping where all the nodes are mapped to each other. They compared three different methods in object recognitions. The experimental results showed that their salient map graph matching is efficient to bipartite matching problem. Besides, the edit distance methodology is used in bipartite graph matching. According to Serratosa's work [29], the bipartite graph matching problem was solved using graph edit distance. In this work, the cost in transforming was computed. In the experiments, the edit distance algorithm took less run time than the bipartite algorithm. The proposed algorithm used the Munkres' algorithm with the cost matrix. The Munkres' algorithm only explored a quadrant of submatrix with the same size of two matrices in comparison. The run time was reduced.

3. THEORETICAL BACKGROUND

In this chapter, the technical methods that are used in the experiments are described. First, the input video sequences are extracted. In this work, the relations between objects in the scene are considered the significant information. To identify the differences between two indoor scenes, the relations between objects are used as input. The indoor scene video contains lines that are used in plane detections. Later planes are detected along with groups of objects lying on those surfaces.

At the present time, the web-based application is widespread, especially video sharing. There are a lot of websites serving any end users to upload video files. Thus the videos in the database are able to be various in size, quality, and content. Usually the videos are classified by input name or tagging. Those appropriate categories were decided by users who uploaded those video clips. Many videos were not assigned description directly as the contents in the videos, so those video clips might be categorized incorrectly by only input name or tagging. Some users preferred to query the video database by using the content such as objects and scene. Giving the meanings to videos helps improving the retrieval. To obtain the meaning of the video, each object in each frame is extracted and labeled. In this research, those extracted objects will become attributes in a node of Vectorized Attributed Graph (VAG).

Nowadays, there are many researches about the video similarity. Their features are separated into low-level and high-level concept. The statistical values in image processing are considered as low-level features. After a frame was segmented into regions creating object-based graph, the connectivity of each object was considered in the process. The connectivity, the relationship among objects in the scene can be considered as high-level features.

TABLE I. COMPARISON OF VIDEO SIMILARITY RESEARCH

Author	Motion Obj. Detection	Background segmentation Type	Low-level features	SIFTor STIP	High-level features	Similarity measurement	Object Annotation
Lee et al. [24] (2005)	✘	Edge	Color, size	✘	RAG	✓	✘
Sun et al. [21] (2011)	Only Vehicle	✘	MFCC,HOG	✓	✘	✘	✓
Cao et al. [10] (2010)	✘	✘	Color, STD	✘	✘	✓	✘
Ding et al. [3] (2010)	✘	Local area	✘	✓	✘	✓	Scene
Zhang et al. [4] (2012)	✘	Local area	MFCC,HOG,HOF	✓	Bag of Event	✓	Event

As shown in the table, using low-level features were limited to small specific object for annotation. For example, the color-based feature is very familiar. Many researches used this type of feature. The disadvantages of color-based feature are the lacking of spatial information and insupportable for different illumination conditions. There are a lot of video similarity researches such as [10] that used only color-based features but adding another technique to overcome those disadvantages. Cao et al. proposed their

video similarity search algorithm. Extracting features of image and video are in the form of Image Characteristic Code (ICC) is based on the statistics of spatial-temporal distribution. Their ICCs were four digits. First three digits were computed from pixel components in YCbCr colors space. The last digit was calculated from the characteristics of Spatial-Temporal Distribution (STD) of image frame feature in video. So their video similarity was related to image frame feature, shot type, length and their temporal variation. Later, the fast search approach for scalable computing was presented based on Clustering Index Table (CIT). Sun et al. proposed Automatic Annotation of Web Videos [21]. They constructed consensus foreground object templates to address moving objects that can be a few kinds of objects. Their extracted features were dense Scale-Invariant Feature Transform (SIFT) and Histogram of Oriented Gradients (HOG). The sparse coding was used to represent both SIFT and HOG for learning the dictionaries and to encode the associated feature descriptors. In addition, they used Mel-frequency spectrum coefficients to extract audio feature. Then Gaussian Mixture Model (GMM) was used to classify the audio. Their categories consisted of vehicle that gave the different sounds such as airplane, ambulance, and fire engine. Furthermore, there are some researches that adapt a technique of bag of visual words for video annotation as well. Ding et al. proposed using emerging patterns weighted with bag of words to increase the performance [3]. SIFT descriptors were extracted and a visual vocabulary was constructed through a k-means clustering algorithm. They limited the input into small local areas before feeding to SIFT extraction by ignoring the relation of the object in those small areas. Later, each cluster was mapped into feature vector or Bag-of-word histogram (BoW histogram). In addition, the mining technique is also helpful to the classification. Finally, the SVM was used to classify those adjusted BoW histograms.

After learning other researches, the result shows that objects are very important in video similarity measures. To compare two scenes in semantic way, further than the containing objects in the scene, their relationships are also significant. Due to the input video are limited to indoor scene only, most of the existing objects are on a plane. From the experiments, single plane scenes were tested as input. Later the results are accurate and have advantage in time consuming compared with other algorithms. However, in the experiment, the input image sequences were simulated with less than ten objects appeared in a scene. Those objects appeared on a plane. In this experiments, detecting object by detecting plane is referred to an extraction method. Finally the input that we need in this work is the set of detected objects with their attributes, describing in next chapter.

3.1. Object on Plane detection

The object detection is the computer process or methods to make the computer understanding that the objects and background are on different layers in single image. In this work, the indoor image are the interested inputs. Normally the objects lie on the surfaces in the upright position. Based on this assumption, the position of camera can be calculated from the distance and sizes of objects on the plane. Thus, first, the plane must be detected. Basically, in indoor scene, the walls and floors are detected from using line detection. According to the referenced work [27], the intersected lines

perform a junction. There are several junction that are defined in the reference. Usually the indoor scene is captured from the inside building views. In nature of man-made, based on geometric structure containing a lot of lines, the appearance of building which can be both inside and outside forms is. These straight lines can be extracted easily by image processing techniques such as Hough transform. Later, the junctions are performed at the point where those lines intersected. Figure 2. shows type of junctions that are considered on the left. On the right, the figure shows an example of preliminary input which is applied by the edge detection and Hough transform.

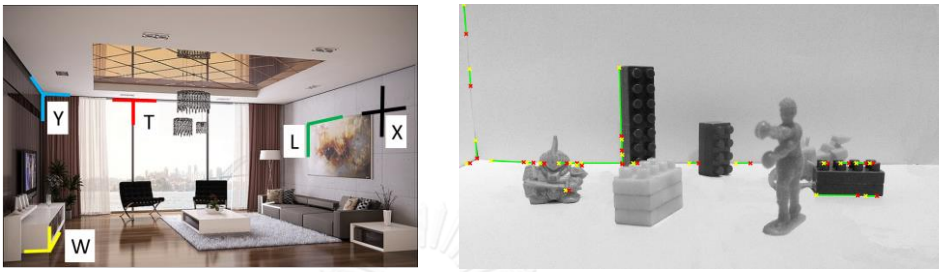


Figure 2. An image of an indoor room with several types of junctions from [27] on the left. On the right, an example of preliminary input before applying edge detection.

Later, the objects are detected. To determine the true objects, the detected plane was used to determine which objects cause false alarm according to the Bao et. al.[26]. In their research, the objects were detected, probably including false alarm. For their object detection, they used detection method with database of objects [30]. To reduce false alarm, they estimated the plane to support the objects' position. They used those objects' positions to determine the plane and improve the detection accuracy. In Figure 3. , the image shows objects on two planes. If all object are detected on the same plane, the meaning is misinterpret. To solve the multi-plane situation, the specific data structure is necessary.

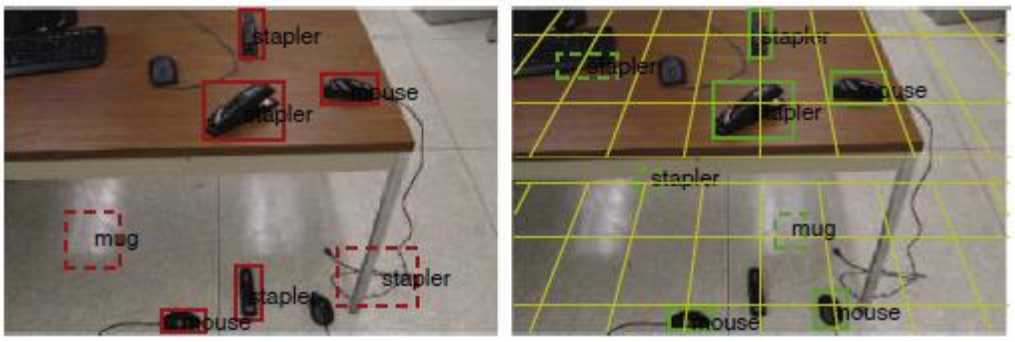


Figure 3. Example scene from [26], from left to right, objects are detected first, and then, two plane are estimated from objects' position.

3.2. Hierarchical graph

After the set of objects and planes existing, in the scene are extracted, those information must be transformed into the form that is easy for people to understand the layout of the scene. The form of data structure that is satisfied the layout of a scene is

now proposed as a graph. The graph that has its nodes as existing objects in the scene is considered. There are various kind of graphs. In this research, the objective of this work is to determine the similarity between two scenes in video sequences. To show the semantic of the scene, a plane also considered as a node. If both objects and planes are nodes, they require specific variable or description to describe those nodes. In figure below, there are two levels of plane in this input scene. According to the detected objects, there are two planes with difference levels which are floor and table. Under table, there are two objects. And there is a monitor on that table. These objects are retrieved from the object detection. This work uses the set of detected objects as input.

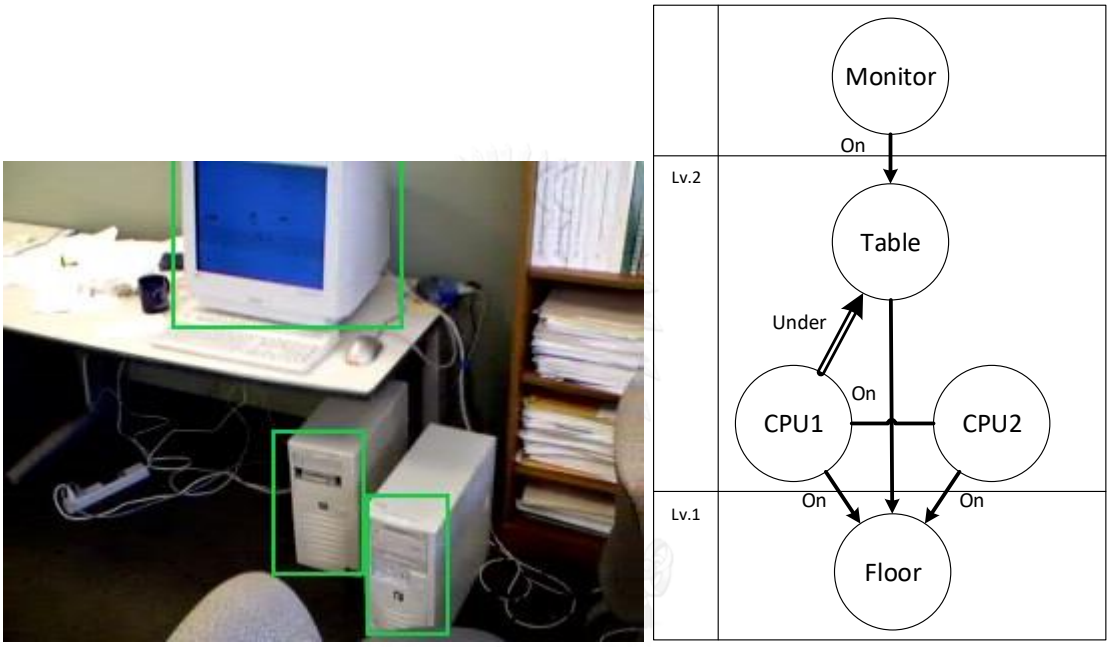


Figure 4. The example of a scene in the dataset with its multi-plane graph.

4. METHODOLOGY

The study in this work started with the single plane simulated scene. To determine the similarity between scenes, the graph is introduced as the structure for presenting a scene. In preliminary experiments, the constraints and assumptions for constructing a graph for a plane are described. The second step is to expand from single to multiple levels of planes. The objects on a plane are applied with single plane as base case. Only there are relationships between planes which are defined further.

The inputs that this work required are detected objects and planes from a scene. Later those objects on each plane are applied with interval-based spatial function. The result from applying interval-based spatial function are the top-view model for each plane. From the top-view model, the vectorized attributed graph (VAG) are created as the representing structure of that scene. The flow of VAG creation is shown as Figure 5. .

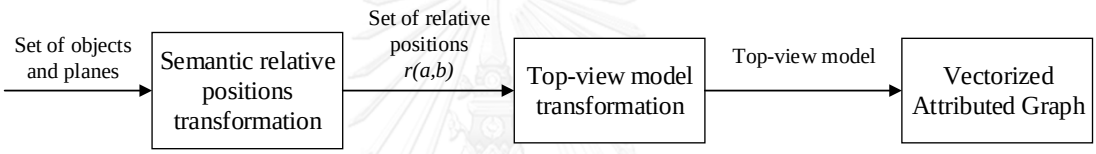


Figure 5. The flow of vectorized attributed graph creation.

4.1. Constraints

In this study, a set of objects is provided as an input for the analysis. There are no assumptions regarding to location of the camera, focus length, and the number of objects in the scene. The studied problem involves a set of objects with attributes, the location of each object, and viewing direction. The definitions of object, location of object, viewing direction, and some other relevant jargon are given as follows.

Definition 1. A projected object (object for short) o_i seen in a 2-dimensional scene from a location l is a 5-tuple of $(x_i; y_i; w_i; h_i; A_i)$ where x_i is the x-coordinate with respect to lower left corner of the image; y_i is the y-coordinate with respect to lower left corner of the image; w_i is projected width; h_i is the projected height; and A_i is a set of attributes of o_i.

$$o_i = (x_i; y_i; w_i; h_i; A_i)$$

where i = objects index.
Note that the $x_i; y_i; w_i; h_i$ are not the actual attribute values of object o_i in the physical world. These coordinates are the relative coordinates with respect to the lower left corner of the scene. The depth information among objects is not provided.

Definition 2. A scene s_i is a 2-dimensional image of size $n_1 \times n_2$ consisting of a set of projected objects.

$$s_i = \{o_1, o_2, o_3, \dots, o_m\}$$

where m is the number of objects.

Definition 3. A viewing direction v_j of a scene s_i is the orthogonal projecting direction of all objects onto s_i with respect to the camera.

Definition 4. Two scenes s_i and s_j are semantically similar if they satisfy the following conditions:

1. The input scenes s_j has the ratio between the summation of objects with the same set of attributes and their attached edges with objects that satisfy with Definition 6 as the comparing scenes and the number of containing objects and edges larger than 60 percent.
2. For any object o_i, the relative positions among o_i and its neighboring objects in both scenes are the same.
3. Both scenes have the same background (wall and ceiling).

Definition 5. There are two *types* of object that is presented in this dissertation. There are possibilities that the objects can lie next to each other. In this case, those objects are considered as a group of objects. In this dissertation, an object and a group of objects are defined with different semantic. Because of our objective is to compare between two scenes in semantic way using object layout. Thus the meanings of a *single object* and a *group of objects* are significant in reducing the time in object comparing. If the object is detected as a *group of objects* meaning that those objects have more than one set of attributes. Surely, the object that is detected as *single object* has a set of attributes. To compare between those different input, there are three possible pairs exist.

1. A *single object* and *single object*: the set of attributes from two objects are compared.
2. A *single object* and a *group of objects*: in this case, two objects are considered as not equivalent.
3. A *group of objects* and a *group of objects*: the attributes of each object in a group is compared with another group. If there is a pair of objects that are not equivalent, both group of object are not equivalent. Both groups are equivalent if objects on both groups are equivalent.

The Figure 6. shows the example of two *types* of objects. In Figure 6. .a, object 2 and 3 are considered as a *group of objects*. This scene will produce a VAG with three nodes. In the other hand, Figure 6. b, object 2 and object 3 are considered as two *single objects* separately. These two scenes have different semantic meaning of objects' layouts.

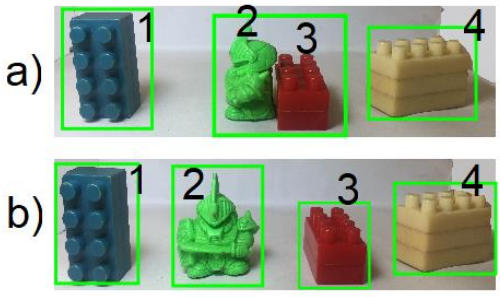


Figure 6. The example of two types of objects.

Definition 6. Two objects o_p and o_q are equivalent if their selected attributes are similar. According to Russell's proposed work, the objects on a scene alignment are detected. Those objects are labelled with semantic by matching with database. The attributes of an object which are used to determine the equivalent are listed below.

1. Semantic, use string matching in comparing this attribute. In the experiments, the string consists of person, robot, motorcycle and box.
2. Color, the HSL is used as an attribute to represent color by separating the lightness. After retrieving the labelled objects, each object is marked with the square area. In the objects, based on segmentation, the largest area of the object is used to represent the color of that object defined by color name. The luminance and shadow in the input images may be various. The degree of Hue along with the range of saturation, the color names are defined as Cyan, Red, Blue-Cyan, Yellow-Red, Blue, Yellow, Magenta-Blue, Green-Yellow, Magenta, Green, Red-Magenta, Cyan-Green, Black and White. The chart of the HSV space is depicted in Appendix.
3. SIFT, or Scale-Invariant Feature Transform, is chosen as an attribute to describe the similarity between two objects that appeared on both scene input. Eventhough the object might be viewed from different angle, their SIFT features are the same for some amounts of points. Thus, in the matching process, the SIFT point matching is used to compare two objects. Two objects are considered as equivalent if SIFT matching is more than 80 percent, due to the different perspective views.

4.2. Assumptions

This work focuses only on the determination of whether two given scenes s_i and s_j with the same set of objects and similar background are exactly the same scene or not. The following situations are assumed.

1. The objects in given scenes s_i and s_j are assumed to be provided prior to our analysis by some image segmentation and object recognition methods such as the one proposed by Bao and Savarese[26].
2. Each object is located on either a floor or another object. The floor is partitioned into grid-based regions.
3. Both considered scenes have the same environmental background. This implies that the background from each scene may not be exactly the same. The background region of an image can be extracted by using line segmentation as introduced in Ramalingam et al is method [27]. For complex background, the approach of Shoaib et al [31] which used key points to estimate the wall, ceiling and floor can be adopted. In addition, the concept of texture classification for detecting wall, ceiling and floor proposed by Hödlmoser and Micusik [32] may also be used.
4. The maximum number of objects in each scene is not fixed.
5. The attributes of an object can be colors, texture, and shape.
6. The edges have their attributes to specify the relation between nodes.
7. The videos are the sequences of images.

4.3. Proposed Concept of Semantic Testing

In this study, the testing of semantic isomorphism of two scenes is based on Definition 4. The difficulty of testing according to Definition 4 is that there is no spatial information of actual relative positions of all projected objects in a 2-dimensional space. The only information provided is the positions of projected objects as seen in the scene. Although some objects may look occluded or overlapped by other objects as they appear on the scene but these objects are not actually occluded or overlapped by the other objects when considering from top view. Hence, satisfying Definition 4 requires the estimated position of each object from top view.

Usually, this testing of similar scenes focuses only on measuring the similar textures appearing in the scenes. But testing similar semantic must focus on the number of appearing objects, the details of objects, their locations, and also the texture of each object. There are several methods for representing temporal knowledge by temporal interval [33] and spatial information by spatial indexing [15] for testing only similar or exact isomorphism between two images or videos. These approaches cannot be used to test whether two scenes with exact number of objects but different shooting angles are semantically similar. However, in this study, the concepts of temporal interval and spatial indexing were modified and extended to cope with testing of semantic similarity. The temporal interval was introduced for describing time of events. Spatial indexing represents the direction and topological relation between objects to retrieve video frames from the video database in a semantic way. Spatial indexing is very useful to summarize the video scene in terms of human language [14] for querying and measuring similarity between scenes in video. To exactly capture the set of objects in a scene and the relative positions of objects, a vectorized attributed graph was introduced in our study. This graph is based on the concept of relative positions in [14] but with various extended modifications to fit our study. The definition of vectorized attributed graph is the following.

Definition 7. A vectorized attributed graph $G_i = (V_i, E_i)$ for scene s_i is an undirected graph consisting of a set of vertices $V_i = \{v_i, \dots, v_m | v_j \text{ represents } o_j \in s_i\}$ and a set of edges $E_i = \{(v_j, v_k) | \text{for some } j \text{ and } k\}$ representing the relative positions among objects derived from top view.

The relative position between two objects will be defined in the next section. Capturing the relative positions among objects in this study differs from how a human determines the similarity of two scenes. Normally with human eyes, people pay more attentions to large objects than small objects so only large objects are considered as landmarks of a scene. Those large objects are served as interested objects. Suppose two scenes with the same object configuration and location are shot from different angles. The effect of different viewing angles may make some objects occlude other objects and cause a wrong interpretation of the relative positions of objects. However, if the object configuration and arrangement are derived from the top view, then the effect of viewing angle can be completely eliminated. The following sequential steps are proposed to test the semantic similarity of two scenes based on Definition 4.

1. In each scene, transforming horizontal relative positions of two close objects in the scene into the logical relative positions in x-coordinates of top view.
2. In each scene, transforming vertical relative positions of two close objects in the scene into the logical relative positions in y-coordinates of top view.
3. In each scene, constructing the VAG graph from all objects and their relative positions in top view.
4. Testing the semantic similarity of two VAG graphs.

4.4. Interval-based Spatial Relation

To estimate the relative positions of two objects in the top view, their relative positions as seen in the scene in both horizontal and vertical directions must be formulated first. The horizontal relative position is formulated by scanning the objects from the left side of the scene. But the vertical relative position is formulated by scanning the objects from the bottom side of the scene upward. Figure 7. shows an example of how to define the relative positions of objects 1 and 2 (denoted by numbers 1 and 2). Object 1 comes before object 2 in horizontal scan and also comes before object 2 in vertical scan. The bottom side of object 1 is closer to the bottom side of the scene than the bottom side of object 2.

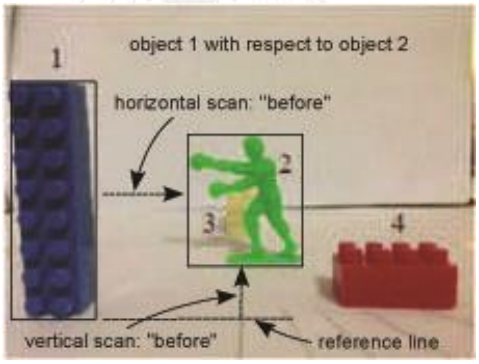


Figure 7. An example of horizontal scan and vertical scan. There are four objects denoted by numbers 1, 2, 3, and 4.

The possible relative positions between two objects in [14] were adapted as a primary consideration. However, new additional relations and names were introduced in this paper to comply with our objectives. In this study, the notations of relative positions of objects a and b in horizontal and vertical scanning directions are written as $f_h\{a,b\}$ and $f_v\{a,b\}$, respectively. $f_h\{a,b\}$ means that the right side of object a is found before the left side of object b in horizontal scan and $f_v\{a,b\}$ means that the bottom side of object a is found before the bottom side of object b in vertical scan. TABLE II. and TABLE III. summarize all possible relative positions of objects a and b as seen in the scene and their $f_h(a,b)$ and $f_v(a,b)$.

TABLE II. is for horizontal scan and TABLE III. is for vertical scan. For example, $f_h(a,b) = (before)$ in the left sub-table means that the right side of object a is found before the left side of object b. But $f_h(a,b) = (contain)$ means that the left side of object a is found before the left side of object b but the right side of object b is found

before the right side of object a. For any objects a and b, their relative positions in both scanning directions are written as

$$r(a,b) = \left(f_h(a,b), f_v(a,b)\right) \qquad (1)$$

As long as, there exist possible relative positions in both horizontal and vertical scanning set. Thus $r(1,2)$ of objects 1 and 2 in Figure 2 is $r(1,2) = (before, before)$. In Table 1, based on what appearing in the scene, the following situations are not covered by the spatial relation between two objects.

1. An object that is on another object or inside another object cannot be recognized. In this case, those objects are expressed as a single object.
2. An object is completely occluded by another larger object. This implies that the occluded object obviously does not appear in the scene.

TABLE II. POSSIBLE INTERVAL-BASED SPATIAL RELATIONS ON LOGICAL GRID, ACCORDING TO SHEARER ET AL. [24] . HORIZONTAL SCAN FROM LEFT TO RIGHT IS PRESENTED. LET A BE THE REFERENCE OBJECT AND B BE THE FELLOWING OBJECT.

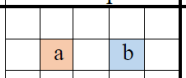
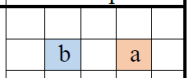
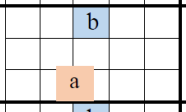
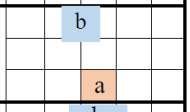
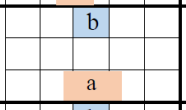
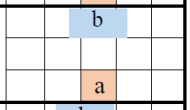
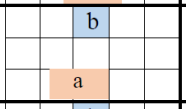
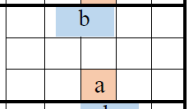
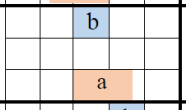
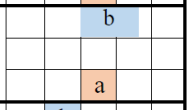
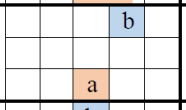
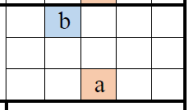
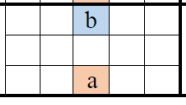
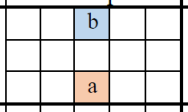
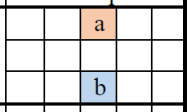
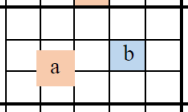
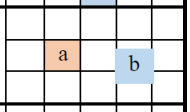
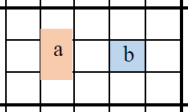
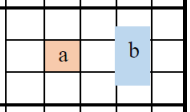
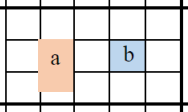
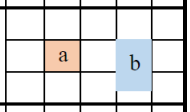
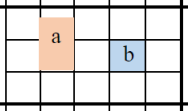
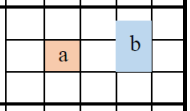
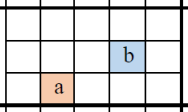
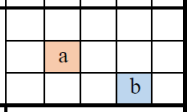
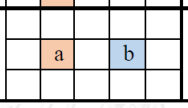
Relationship	Example	Relationship	Example
$f_h(a,b) = (before)$		$f_h(a,b) = (after)$	
$f_h(a,b) = (overlap)$		$f_h(a,b) = (overlap\ invese)$	
$f_h(a,b) = (contain)$		$f_h(a,b) = (contain\ inverse)$	
$f_h(a,b) = (align - R)$		$f_h(a,b) = (align - R\ inverse)$	
$f_h(a,b) = (align - L)$		$f_h(a,b) = (align - L\ inverse)$	
$f_h(a,b) = (meet)$		$f_h(a,b) = (meet\ inverse)$	
$f_h(a,b) = (equal)$			

TABLE III. POSSIBLE INTERVAL-BASED SPATIAL RELATIONS ON LOGICAL GRID, ACCORDING TO SHEARER ET AL. [24]. VERTICAL SCAN FROM THE CLOSEST TO THE DEEPEST OF SHOOTING CAMERA IS PRESENTED. LET A BE THE REFERENCE OBJECT AND B BE THE FELLOWING OBJECT.

Relationship	Example	Relationship	Example
$f_v(a,b) = (before)$		$f_v(a,b) = (after)$	
$f_v(a,b) = (overlap)$		$f_v(a,b) = (overlap\ inverse)$	
$f_v(a,b) = (contain)$		$f_v(a,b) = (contain\ inverse)$	
$f_v(a,b) = (align - T)$		$f_v(a,b) = (align - T\ inverse)$	
$f_v(a,b) = (align - B)$		$f_v(a,b) = (align - B\ inverse)$	
$f_v(a,b) = (meet)$		$f_v(a,b) = (meet\ inverse)$	
$f_v(a,b) = (equal)$			

The information of spatial relation of objects from both horizontal and vertical scans will be used to sketch the logical positions of the objects as appearing in the top view. Note that since the actual depth information of all objects cannot be concluded from the given scene, the top-viewed positions of all objects must be only logical not physical. The concept of how top-viewed positions are sketched is in the next section.

4.5. Sketching Top-viewed Positions of objects

The top-viewed plane is logically partitioned into a set of one-unit squares with the origin at the lower left corner. Since the depth of each object is unknown from the scene, so the width is the only useful information for representing the size of object. This width when appearing in the top view lies along the x-axis of the top-viewed plane. To completely define the size of each object in the top view, the length of object in the y-axis is assumed to be one unit length. Figure 8. shows the structure of top-viewed plane and an example of top-viewed size of object.

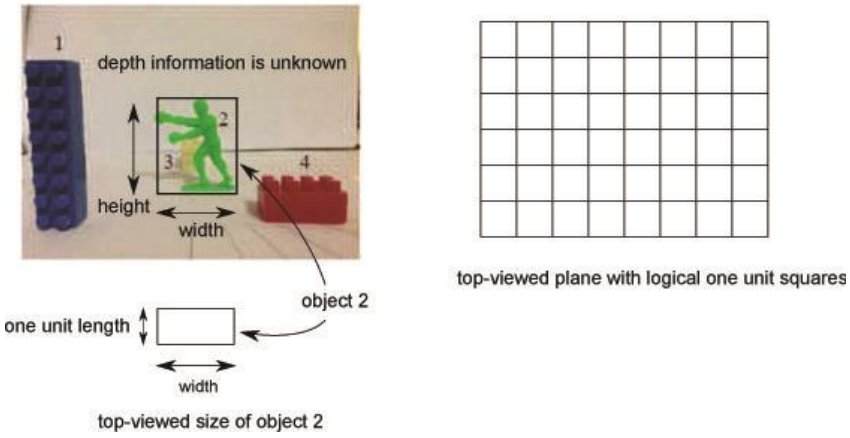


Figure 8. Logical grid structure of top-viewed plane and an example of top-viewed size of object 2.

Sketching the positions of objects in top view consists of two main steps. The first step is collecting a set of object relations based on the relations shown in Table 1. The second step is to assign logical position to each object defined by the relations. The details both steps are in Algorithm 1 and Algorithm 2.

Algorithm 1: Collecting pairs of relative objects from the scene

Input: $s_a = \{o_1, o_2, o_3, \ldots, o_m\}$ with their unit sizes.
Output: All relevant pairs of relative objects.
1. Put a rectangle to cover each object oi.
2. Let $H = \emptyset$ be the set of horizontal relative objects.
3. Let $V = \emptyset$ be the set of vertical relative objects.
4. Let L_j be a set of objects whose left sides are found in the j^{th} order of scanning sequence.
5. Horizontally scan the scene from the left side towards the right side.
6. **For each** o_i whose left side is found at position j do
7. $L_j = L_j \cup \{o_i\}$.
8. **EndFor**
9. **For each** pair $L_j$and L_{j+1} do
10. Pair up (o_i, o_k) such that $o_i \in L_j$ and $o_k \in L_{j+1}$.
11. Set $H = H \cup \{(o_i, o_k)\}$.
12. **If** L_j whose $|L_j| > 1$ **then**
13. Pair up all (o_i, o_k) such that $o_i, o_k \in L_j$ and $o_i \neq o_k$.
14. Set $H = H \cup \{\forall(o_i, o_k)\}$.
15. **EndIf**
16. **EndFor**
17. Let B_j be a set of objects whose bottom sides are found in the j^{th} order of scanning sequence.
18. Vertically scan the scene from the bottom side upwards the top side.
19. **For each** o_i whose bottom side is found at position j do
20. $B_j = B_j \cup \{o_i\}$ = Bj.
21. **EndFor**
22. **For each** pair B_j and B_{j+1} do
23. Pair up (o_i, o_k) such that oi 2 $o_i \in B_j$ and $o_k \in B_{j+1}$.
24. Set $V = V \cup \{(o_i, o_k)\}$.
25. **If** B_j whose $|B_j| > 1$ **then**
26. Pair up all (o_i, o_k) such that $o_i, o_k \in B_j$and $o_i \neq o_k$.

```
27.          Set V = V ∪ {∀(o_i, o_k)}.
28.    EndIf
29. EndFor
```

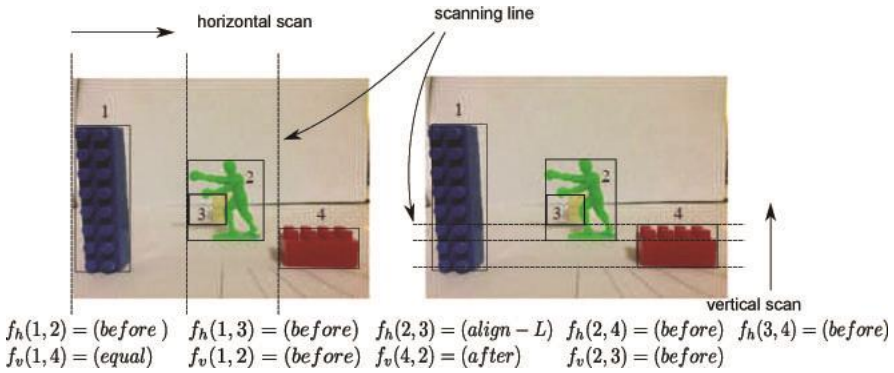


Figure 9. An example of how to find pairs of relative objects. There are four rectangles covering objects. The left sub-image shows the horizontal scan and the right sub-image shows the vertical scan.

Figure 9. shows an example of how to find all possible relative pairs. From the horizontal scan, the objects in each L_i are $L_1 = \{1\}; L_2 = \{2,3\}; L_3 = \{4\}$. Since L_2 has two objects, objects 2 and 3 must be paired up as $(2,3)$. Hence, all pairs obtained after 7 step 16 are $H = \{(1,2), (1,3), (2,3), (2,4), (3,4)\}$. Similarly, from the vertical scan, the objects in each B_i are $B_1 = \{1,4\}; B_2 = \{2\}; B_3 = \{3\}$. Note that B1 has two objects. So these two objects must be paired up as $(1,4)$. All pairs obtained after step 29 are $V = \{(1,4), (1,2), (4,2), (2,3)\}$.

After set H and V are obtained. The relationship for each pair of objects in set H and V are represented by interval-based spatial relation. For example, for $H, f_h(1,2) = (before), f_h(2,3) = (align - L), f_h(2,4) = (before), f_h(3,4) = (before)$, as shown in Figure 10. The examples of V are shown in Figure 11. After obtaining all possible pairs from horizontal and vertical scan, the logical top-viewed positions of each relative object pair (a, b) are defined by considering $f_h(a, b)$ and $f_v(a, b)$. The detail is presented in Algorithm 2 and Algorithm 3.

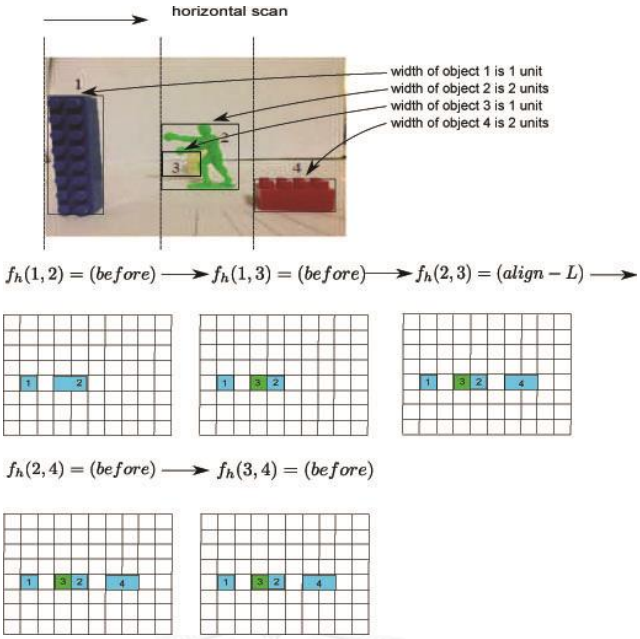


Figure 10. An example of how to logically position each relative object pair in set H.

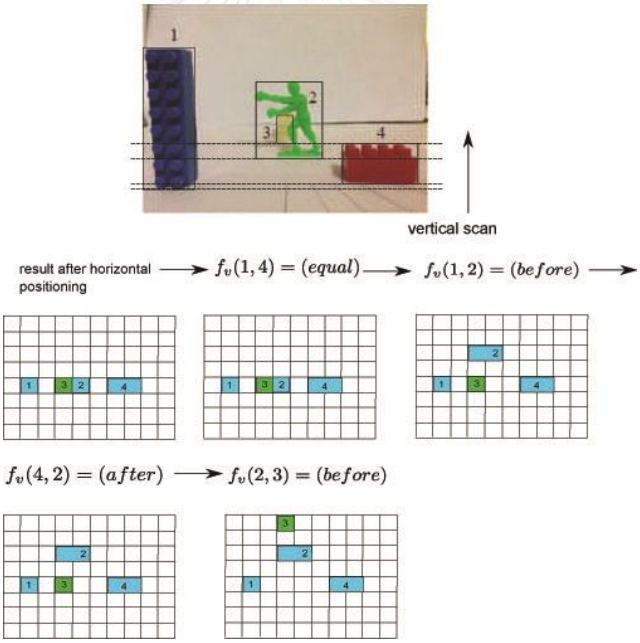


Figure 11. An example of how to logically position each relative object pair in set V.

Figure 10. and Figure 11. show the sequence of placing each object found in horizontal scan and vertical scan. The horizontal positions are defined before the vertical positions according to the finding sequence of each object. The coordinates of each object are defined by the relations in TABLE II. and TABLE III.

Algorithm 2: Horizontal positioning each object in the top-viewed plane

Input: Set H.
Output: Horizontal relative positions of all objects on top-viewed plane.
1. Set the unit width w_i and unit length l_i a to each object o_i.

2. Randomly select an initial position on the top-viewed plane for the first object.
3. **For each** pair $(o_i, o_j) \in H$ do
4. **Case** $f_h(o_i, o_j)$:
5. *before*: position o_i at coordinate (x_i, y_i) and o_j at $(x_i + w_i + 1, y_i)$.
6. *overlap*: position o_i at coordinate (x_i, y_i) and o_j at $(x_i + 0.5w_i, y_i)$.
7. *contain*: position o_i at coordinate (x_i, y_i) and o_j at $(x_i + 0.5w_i, y_i)$.
8. *align-R*: position o_i at coordinate (x_i, y_i) and o_j at $(x_i + w_i - w_j, y_i)$.
9. *align-L*: position o_i at coordinate (x_i, y_i) and o_j at (x_i, y_i).
10. *meet*: position o_i at coordinate (x_i, y_i) and o_j at$(x_i + w_i, y_i)$.
11. *equal*: position o_i at coordinate (x_i, y_i) and o_j at (x_i, y_i).
12. *after*: position o_i at coordinate (x_i, y_i) and o_j at $(x_i - w_j - 1, y_i)$.
13. *overlap* inverse: position o_i at coordinate (x_i, y_i) and o_j at $(x_i - 0.5w_j, y_i)$.
14. *contain inverse*: position o_i at coordinate (x_i, y_i) and o_j at $(x_i - 0.5w_i, y_i)$.
15. *align-R inverse*: position o_i at coordinate (x_i, y_i) and o_j at $(x_i + w_i - w_j, y_i)$.
16. *align-L inverse*: position o_i at coordinate (x_i, y_i) and o_j at (x_i, y_i).
17. *meet inverse*: position o_i at coordinate (x_i, y_i) and o_j at $(x_i - w_j, y_i)$.
18. **EndCase**
19. **EndFor**

Algorithm 3: Vertical positioning each object in the top-viewed plane

Input: Set V.
Output: Vertical relative positions of all objects on top-viewed plane.
1. Set the unit width w_i and unit length l_i to each object o_i.
2. **For each** pair $(o_i, o_j) \in V$ do
3. **Case** $f_v(o_i, o_j)$:
4. *before*: position o_i at coordinate (x_i, y_i) and o_j at $(x_j, y_i + l_i + 1)$.
5. *overlap*: position o_i at coordinate (x_i, y_i) and o_j at $(x_j, y_i + 0.5)$.
6. *contain*: position o_i at coordinate (x_i, y_i) and o_j at $(x_j, y_i + 0.5)$.
7. *align-T*: position o_i at coordinate (x_i, y_i) and o_j at $(x_j, y_i + l_i + l_j)$.
8. *align-B*: position o_i at coordinate (x_i, y_i) and o_j at (x_j, y_i).
9. *meet*: position o_i at coordinate (x_i, y_i) and o_j at $(x_j, y_i + l_i)$.
10. *equal*: position o_i at coordinate (x_i, y_i) and o_j at (x_j, y_i).
11. *after*: position o_i at coordinate (x_i, y_i) and o_j at $(x_j, y_i + l_i + 1)$.
12. *overlap inverse*: position o_i at coordinate (x_i, y_i) and o_j at $(x_j, y_i - 0.5)$.
13. *contain inverse*: position o_i at coordinate (x_i, y_i) and o_j at $(x_j, y_i - 0.5)$.
14. *align-T inverse*: position o_i at coordinate (x_i, y_i) and o_j at $(x_j, y_i - 0.5)$.
15. *align-B inverse*: position o_i at coordinate (x_i, y_i) and o_j at (x_j, y_i).
16. *meet inverse*: position o_i at coordinate (x_i, y_i) and o_j at $(x_j, y_i + l_j)$.
17. **EndCase**
18. **EndFor**

4.6. Constructing Vectorized Attributed Graph from Top-Viewed Plane

The relative positions of all objects positioned on the top-viewed plane can be captured for the semantic isomorphism analysis by using a semantic spatial graph defined in Definition 5. Actually, the logical positions in the horizontal direction from Algorithm 2 and the logical positions in the vertical direction from Algorithm 3 already imply the relative positions among all nearest objects. From both algorithms, the positioning order of each object and its logical coordinates in horizontal and vertical directions are obtained. After all objects in the scene are paired with their relationships by vertical and horizontal scanning, the sets H and V show all closest pairs in both directions. The relative positions of objects in the scene are defined by intersected items between set H and V, according to (1). From the Figure 7. , the relative positions of each pair are consisted as following list:

1. $r(1,2) = (before, before)$
2. $r(2,4) = (before, after)$
3. $r(2,3) = (align - L, before)$
4. $r(1,4) = (before, equal)$

A Vectorized Attributed graph in the experiment is represented by an adjacency matrix. As mention about the relative positions $r(a, b)$, each of them shows the closest relation between two objects. Therefore, an edge in Vectorized Attributed graph is performed according to a relative position. Constructing a vectorized attributed graph can be simply started from the result of Algorithm 2 followed by the result of Algorithm 3 as detailed in Algorithm 4.

The relative position from the same scene but different perspective might different but their adjacency matrices are still isomorphic. For example from the scene in figure 9, its relative positions are mentioned above. If this scene is viewed in another different for 180°, its 180° different angle perspective defines different relative position function. Its relative position functions are consisted as following list:

1. $r(1,2) = (after, after)$
2. $r(2,4) = (after, after)$
3. $r(2,3) = (align - R, after)$
4. $r(1,4) = (after, equal)$

Eventhough both relation spatial function are not the same, the objects of closest pair that is selected are the same. Thus semantic spatial matrix is assigned as a symmetric matrix.

Constructing a vectorized attributed graph can be simply started from the result of Algorithm 2 followed by the result of Algorithm 3 as detailed in Algorithm 4.

Algorithm 4: Constructing Vectorized Attributed graph in Matrix Form

Input: Sets of $r(o_i, o_j)$.
Output: A matrix capturing all relevant relative objects in top-viewed plane.
1. Let M be a zero matrix of size $n \times n$, where n is the number of objects in the scene.
2. **For each** pair $r(o_i, o_j)$ do
3. $M(i,j) = 1$

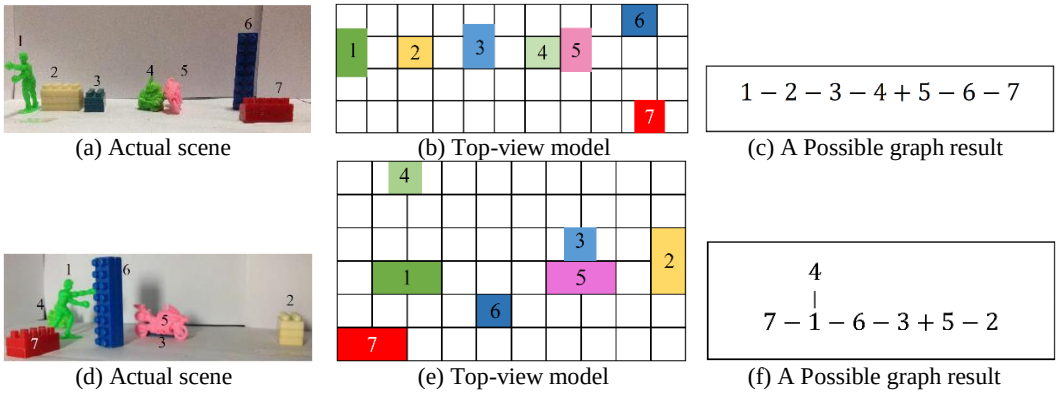
4. $M(j,i) = 1$
5. **EndFor**

4.7. The edge between objects

The relationships that perform edges in the set E, for each G present the distances between any two closest objects. There might be some restrictions that can be occurred and caused the confusion. The restrictions are described as follows.

1. The edge in semantic spatial graph represents the nearest reachable objects. For example as Figure 7. , $r(1,3)$ cannot be included in the set of r. This edge is excluded because there is an object in-between them in vertical distance, even though the $r(1,3)$ is in set H. Besides, the relation between object 1 and 2 is in both H and V, the edge between them is performed as $r(1,2)$. It is represented as $e(1,2) = 1 - 2$ in vectorized attributed graph. In Figure 7. , object 1 performs connections with object 2
$$E = \{e(1,2)\},$$
where $e(1,2) = 1 - 2$.
2. Objects that are placed next to each other are represented as one vertex in the semantic spatial graph. That is,
$$\forall f_h \forall f_v (r(a,b) = (f_h, f_v) \wedge (f_h \vee f_v) \in \{meet, meet\ inverse\} \wedge (f_h, f_v) \notin \{before, after\}) \rightarrow r(a,b) = o_{a+b}$$
where $r(a,b)$ is spatial relation position, a and b are indices of objects and f_h, f_v are sets of possible interval-based spatial relation for horizontal and vertical scan. For example in figure 8, $r(3,5) = o_{3+5}$. Those two objects are considered as one object.
3. By selecting nearest objects with incident edges connected, a semantic spatial graph is performed. Hence there is no object without an edge.

Vectorized attributed graph comparison is introduced. Because of the number of objects is not large, the comparing is quite not complicated. By comparing the edges that incident to each vertex containing the same attributes, only two edges are compare for each vertex.



(a) Actual scene (b) Top-view model (c) A Possible graph result

(d) Actual scene (e) Top-view model (f) A Possible graph result

Figure 12. The example of two different video scenes (a) and (d) which are transform to top-view model in (b) and (e). (c) and (f) are the extracted VAG from (b) and (e) accordingly.

The Vectorized Attributed Graph of one scene is isomorphic to another scene. In case, they satisfy conditions as in Definition 4. It can be noted that those scenes are similar.

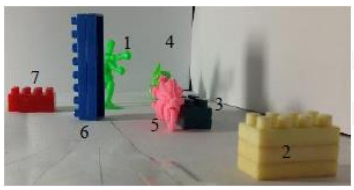
$$Iff\ G_I \equiv G_J \Rightarrow I \equiv J$$

where I and J are different scenes.

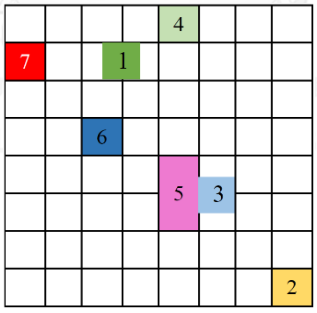
To compare similarity between scenes, their semantic spatial graph patterns are considered as the main features in comparison. The vertices are compared after mapping the similar objects in both scenes into vertices. If their semantic spatial graphs that represent two scenes are exactly the same, they are considered as having the same background. For proving this theorem, the same background with different angles of view are compared. Even though the camera captures the background in different angle, their vectorized attributed graphs are similar and two scenes are considered as equivalent. Figure 13. shows the different angle from the same background of Figure 12. d. Their semantic spatial graphs are similar.

Degrees, number of edges and number of vertices are used to determine an isomorphism between two graphs if sub-cycles do not exist on both graphs.

According to section 4.4, the construction of graph prevents the selected objects to be selected for connecting nearest objects together. It is impossible that a semantic spatial graph has a sub-cycle. The structure of graph is least complex. Therefore only their adjacency matrices can be used to determine whether they are isomorphic. For example, an adjacency matrix from scene in Figure 19. is the same as that in Figure 13. for different angles.



(a) Actual scene



(b) Top-view model

4

$7-1-6-3+5-2$

(c) A possible semantic spatial graph

	o_1	o_2	o_3	o_4	o_5	o_6
o_1	0	0	0	1	1	1
o_2	0	0	1	0	0	0
o_3	0	1	0	0	1	0
o_4	1	0	0	0	0	0
o_5	1	0	1	0	0	0
o_6	1	0	0	0	0	0

(d) An adjacency matrix

Figure 13. shows a same video with Figure 12. but in different perspective, along with its top-view model and semantic spatial graph.

4.8. Graph Comparison

People recognize scenes by the objects and surrounding. By looking at objects, they determined the similarity using their relationship such as distances and positions. In this paper, semantic spatial graph is introduced. A node represents an object. By considering their nearest objects, an edge is found between nodes to present the relationship which

is the closeness between objects. The graph comparison methodology is described below.

1. After a graph is obtained, each graph is in a form of an adjacency matrix as input for comparison.
2. The matrix dimension is m^2 where m is number of objects in that graph.
3. The group of objects that are next to each other is represented as an object. This object has special attributes as group of objects, as described in section the edge between objects. In Figure 13. the adjacency matrix shows only six objects. o_{3+5} represents the object number 3 and 5 that are next to each other. o_3 and o_5 represent objects numbered 3 and 5 accordingly.
4. The number 1 in an adjacency matrix means there is an edge between objects of that row and column. Certainly, diagonal entries of an adjacency matrix are zeros. Besides, this matrix contains either one or zero.
5. If their numbers of objects are equal, their objects attributes are then compared individually with each other. There may be some objects but not all appearing on both scenes. This case is considered as not isomorphic.

The semantic spatial graph comparison algorithm is shown in Algorithm 5.

Algorithm 5: Comparing Semantic Spatial Graph in Adjacency Matrix Form, in case $m \geq n$.

```
Input: Adjacency matrix M1 and M2 of graph G1 and G2, accordingly.
Output: Similarity score, maximum is equal to 1.
1. Let M1  be an adjacency matrix of graph G1, m × m.
2. Let M2 be an adjacency matrix of graph G2, n × n.
3. Let O1 be a set of objects in G1. Let nEdge = 0.
4. Let O2 be a set of objects in G2.
5. Let matchM be a matrix that keeps pairs of matched objects between
O1  and O2.
6. For each m do
7.    For each n do
8.       Compare each object in both scenes.
9.       If types of objects in O1 and O2 are group of objects then
10.         Compare all sets of attributes in both groups.
11.         If all objects in both groups of O1 and O2 are
                 equivalent according to Definition 5 and 6 then
12.            Store the matched object of O2 into matchM ordered as
               O1 index
13.         EndIf
14.      EndIf
15.      If types of objects in O1 and O2 are single object then
16.         Compare the set of attributes from both objects.
17.         If the set of attributes from objects in O1 and O2 are
                 equivalent according to Definition 5 and 6 then
18.            Store the matched object of O2 into matchM ordered as
               O1 index.
19.         EndIf
20.      EndIf
21.   EndFor
22. EndFor
23. For each matchM do
24.   If the element of row and column in M1 and M2
```

```
25.          according to matchM are equal. then
26.       increase nEdge by 1
27.    EndIf
28. EndFor
29. Return (nEdge + |matchM|)/(|M1| + |O1|)
```

	o_1	o_2	o_3	o_4	o_5	o_6	o_7
o_1	0	0	0	1	1	0	0
o_2	0	0	0	0	0	0	1
o_3	0	0	0	0	1	0	0
o_4	1	0	0	0	0	1	0
o_5	1	0	1	0	0	0	0
o_6	0	0	0	1	0	0	1
o_7	0	1	0	0	0	1	0

(a) Adjacency matrix of top scene from figure 19

	o_1	o_2	o_3	o_4	o_5	o_6	o_7
o_1	0	0	0	1	1	0	0
o_2	0	0	0	0	0	0	1
o_3	0	0	0	0	1	0	0
o_4	1	0	0	0	0	1	0
o_5	1	0	1	0	0	0	0
o_6	0	0	0	1	0	0	1
o_7	0	1	0	0	0	1	0

(b) Adjacency matrix of middle scene from figure 19

Figure 14. shows adjacency matrices of figure 19, top and middle figures accordingly.

For comparing semantic spatial graph, every two nodes of two graphs are compared to identify the same object that appeared on both graphs. Each node has attributes describing itself such as color, texture, material and shape. If n is the number of objects, the adjacency matrix of a graph is as $n \times n$ matrix. In Figure 14. , there are examples of adjacency matrices which constructed from figure 19 top and middle pictures. Assuming that both graphs are not much different in number of objects, time complexity for the object comparison is $O(n^2)$. After objects are matched, each matched pair is compared to the attached edges and their destinations. In this process, time complexity is $O(n^2)$ for the worst case. The graph matching in the proposed algorithm is similar to matrices comparison. However the structure of semantic spatial graph is not dense as complete graph, in reality, the possibility that graph including n nodes with $n-1$ edges attached is low or this case barely occurred.

4.9. The video sequence comparison

There exist two video sequences. To find the similar sequences in those two videos, frames in both videos are compared using the proposed method. In this work, the proposed method has the advantages in comparing between two scenes using the relationship between objects. Those relationships perform graphs of the objects in which a node represents an object while an edge represents their relationship. In the experiments, video sequences of indoor scenes are captured from many rooms. Each video sequence contains no moving objects but static objects. The video camera during the recording was swiveling. A frame will contain some of objects from contiguous frames. When comparing between contiguous frames, their similarity rate might slightly different due to the swiveling of the camera during the recording.

Definition 8. Define video input as D. Video sequence D contains frames F.

$$D = \{F_1, F_2, \dots, F_i\} \text{ where } i = \text{number of frames.}$$

If a graph represents a frame, for a video D, there are i graphs which are produced. The input video sequence is analyzed using each frame to compare with their contiguous frames. The ratio of their similarity might not be much different. Figure 15.

shows the processing of comparison within the sequence of video. The similarity ratio shows the percentage of the similarity of existing objects and their arrangements with the compared one. The sequence of ratio values between each frame is used to show the smoothness of that video sequence. If there exists a change of scene in that video, the histogram of ratio will give a low value of the ratio sequence. Figure 16. shows the method in comparing between two sequences of videos. Each frame in both sequences are compared, one on one.

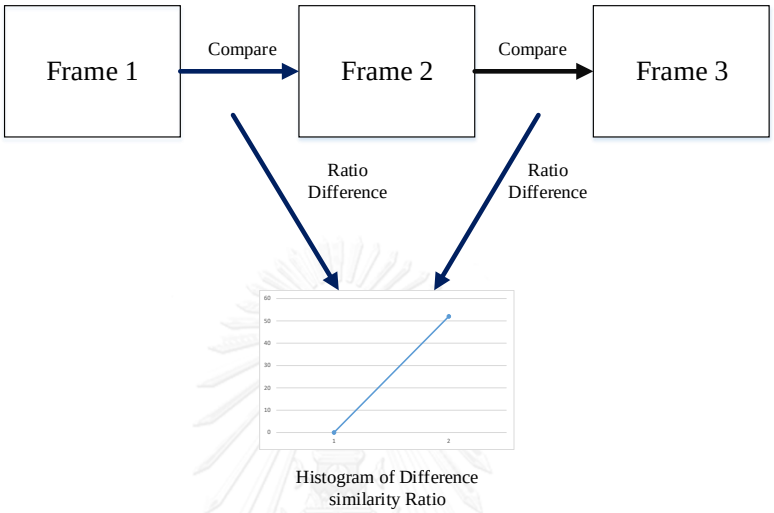


Figure 15. Method in comparing in the same sequence of video.

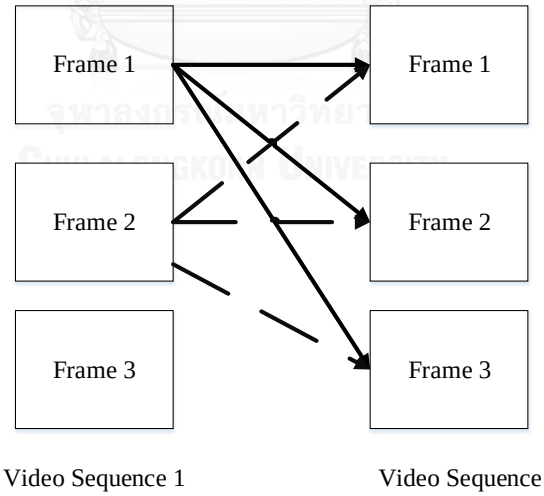


Figure 16. Method in comparing between two sequences of videos.

In summary, the graph comparison between semantic spatial graphs has time complexity of $O(n^2)$, in case both graphs have their numbers of objects equal to n or not much different. The performance of the proposed algorithm is proved by examples in the next chapter.

6. EXPERIMENTAL RESULTS

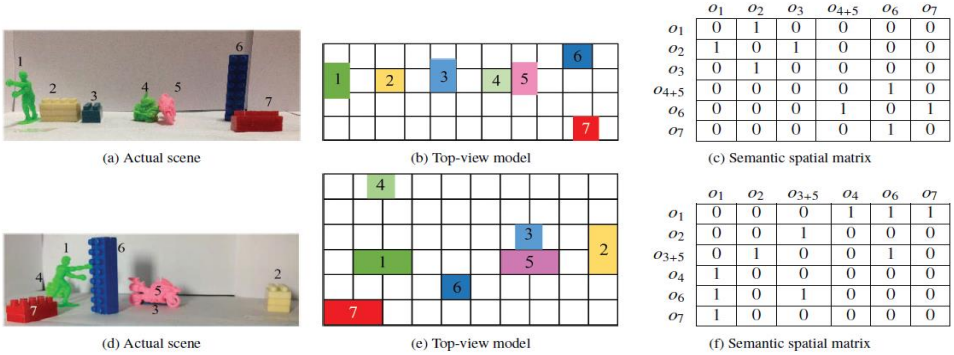
In this work, the experiments are separated into two parts. The first part is to prove that a graph is satisfied to represent a scene. The single plane scenes are simulated as input. Second part, the video sequences of indoor scene are used as the dataset. The graph to represent multi-plane is also introduced. As the graph is introduced to be used in this experiment, other graph methods for graph similarity are used to evaluate with proposed work.

6.1. Evaluation

In the experiments, there are two algorithms that are used to compare with proposed method. The algorithm in [14] are used to compare with the proposed algorithm. In their experiment, they use a frame of video to build decision tree model. Later they used that tree to find the indexing by comparing with other frames. During creating step of decision tree model, they have to do the permutation for constructing the graph and during comparison they have to find the permutation of compared graph to find the best path. In the other hands, they use their algorithm to compare with A* star search. A* search is the best first search by the algorithm which will find the best path to the solution. Thus they compare time complexity between A* search and their decision tree largest common subgraph. Besides, in the experiment, the similarity rate are computed to evaluate the percentage of similarity with the existing objects and their arrangements between two VAG graphs using the below expression.

$$R = \frac{(number\ of\ equivalent\ nodes\ in\ s_i\ and\ s_j + number\ of\ edges\ attached\ with\ equivalent\ nodes)}{number\ of\ objects\ and\ edges\ in\ s_j}$$

For example in the Figure 17. , there are 7 objects that are equivalent from comparing objects in those two scenes. But their arrangements are not even close to each other. According to the definition 5, the group of objects in those two graphs are not equivalent though the object inside the group are equivalent. The similarity rate for scene a to scene b is $\frac{5+0}{10}$. The rate is 0.5. Because of the adjacency matrices are symmetric for single plane experiment, the number of edges are divided by two. The



	o_1	o_2	o_3	o_{4+5}	o_6	o_7
o_1	0	1	0	0	0	0
o_2	1	0	1	0	0	0
o_3	0	1	0	0	0	0
o_{4+5}	0	0	0	0	1	0
o_6	0	0	0	1	0	1
o_7	0	0	0	0	1	0

	o_1	o_2	o_{3+5}	o_4	o_6	o_7
o_1	0	0	0	1	1	1
o_2	0	0	1	0	0	0
o_{3+5}	0	1	0	0	1	0
o_4	1	0	0	0	0	0
o_6	1	0	1	0	0	0
o_7	1	0	0	0	0	0

Figure 17. The examples of two scenes with their top-view model and adjacency matrices.

same as scene a, scene b has the similarity rate as $\frac{5+0}{11}$. The rate is 0.45. This interprets that they are almost similar in which at least the objects in both scene are equivalent to each other.

6.2. The Experiment Single Plane

To emphasize the theories, some examples were reported in this section along with the discussion. According to the hypothesis, though the video background was captured from different angle, those video background top-view models will be the same. In the experiment, seven objects were placed on a plain. Those formations acted as a scene. Ten backgrounds were simulated. A background was captured with ten angles for ten scenes. The camera faced a wall perpendicularly and move around counterclockwise to face another wall perpendicularly. There were ten degrees ranging from 0° to 90° degree of capturing angles with step size of 10 degree. This experiment can be expressed in definition 3, as follows.

$$s_i = \{v_1, v_2, \dots, v_{10}\} \; where \; i = 1,2, \dots ,10$$

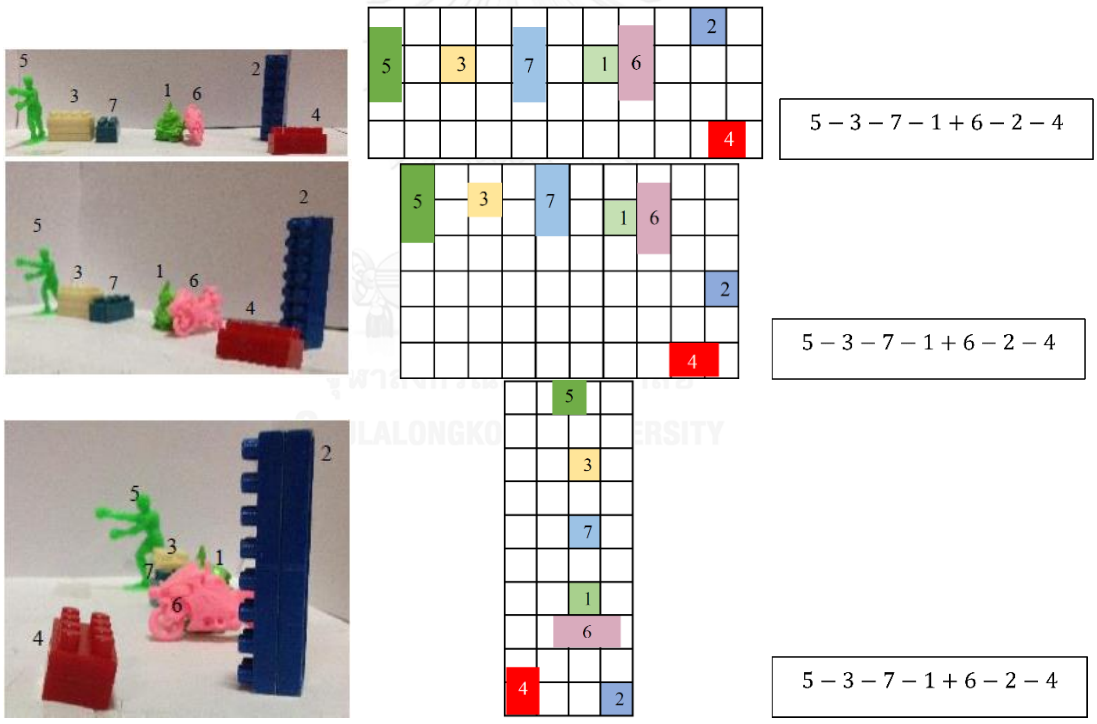


Figure 18. From top to bottom, the example of scene number 3 in database, a scene was captured with different angle of 0°, 40°, and 90° to obtain three backgrounds. From left to right, background image of a scene, its top-view model, and semantic spatial graph.

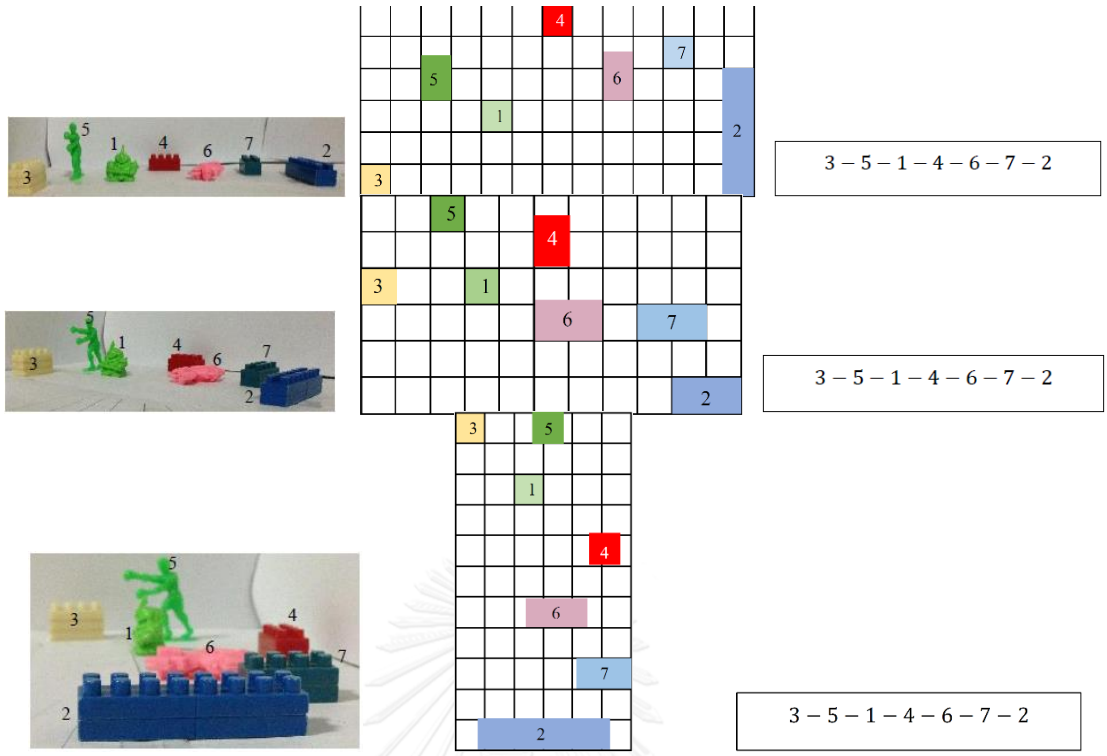


Figure 19. From top to bottom, the example of scene number 5 in database, a scene was captured with different angle of 0°, 40°, and 90° to obtain three backgrounds. From left to right, background image of a scene, its top-view model, and semantic spatial graph.

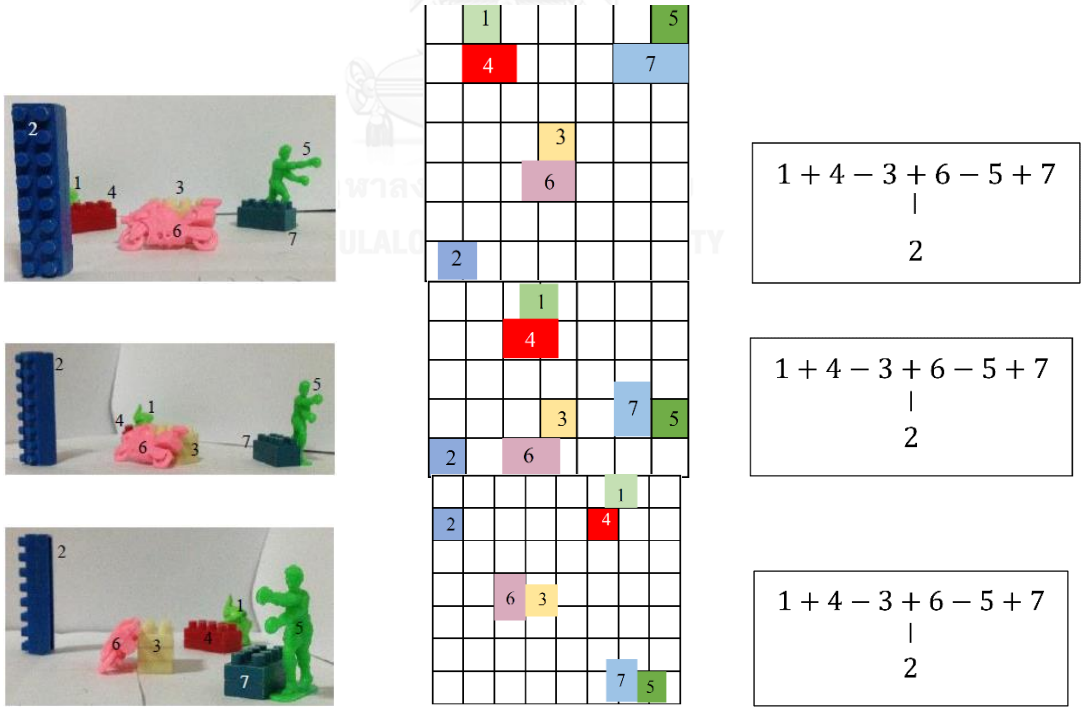


Figure 20. From top to bottom, the example of scene number 9 in database, a scene was captured with different angle of 0°, 40°, and 90° to obtain three backgrounds. From left to right, background image of a scene, its top-view model, and semantic spatial graph.

Figure 18. , Figure 19. , Figure 20. show images which were captured from three different angles of the same scene. Each scene provided semantic spatial graph. The above definitions were applied to these examples below. The results indicated objects' positions for each scene. There is possibility that top-view model affects the proposed algorithm to become effective. If the top-view is exactly the same as its scene, the semantic spatial graph is efficient to be used to compare similarity between scenes. In conclusion, one hundred pictures were used in this experiment to prove the performance of semantic spatial graph similarity via top-view model algorithm. Each top-view model produces a semantic spatial graph. Each graph has at most seven vertices due to the seven objects in the background set up.

The experiments were separated into accuracy and performance measurements. A* and decision tree[14] were implemented for comparison of time complexity. The result will prove that proposed graph structures with proposed graph matching used least time consuming though all accuracies were not much different.

A* and decision tree need tree model for comparing with a sequence. In A* algorithm, all nodes are calculated to find the best sequence as the input. Decision tree was introduced by Shearer et al.[14]. Their algorithm is used to find largest common subgraph. The algorithm need prior knowledge but they are good for video indexing and retrieval. In their experiments, decision tree LCSG algorithm is faster than A* algorithm. In their matching algorithm, they do the permutation for input query until they matched with all existed objects. Therefore their algorithm takes less time than A* algorithm.

6.3. The Experimental Result of Single Plane in term of Accuracy

After the semantic spatial graph for a scene were produced, it was used to determine the similarity between two scenes using graph comparison. TABLE IV. - TABLE VI. show the ratio of scene 7 which was captured with 0° perspective views. The ratios were compared with those of three different algorithms, that is A*, Decision tree with largest common sub-graph and proposed algorithm. The other results that are the ratio of comparing between image model and the dataset for all algorithms are shown in the Appendix. Figure 21. shows ten scenes as the examples of dataset. Their top-view models and vectorized attributed graphs for various perspective views are the same.

Objects' formation is performed in many situations under indoor and outdoor environments. For example circle formation is simulated from meeting room and star formation is simulated from dining room and living room. Some rooms in a same type may be different according to their furnishing. From TABLE IV. , it is possible that same scene with different perspective provides different semantic spatial graph. The causes of differences are hidden objects and wrong determining objects' positions. For hidden objects, when a scene is seen in different perspective views, some objects might be occluded by other objects. This will cause the semantic spatial graph not isomorphic with the others produced from the same scene. Though the graph are not isomorphic, their subgraphs might be isomorphic. Wrong determining the objects' positions causes

wrong logical functions leading to false interpreting to top-view model. Either problems can be prevented using subgraph isomorphic instead.

TABLE IV. SHOWS THE RESULT OF SIMIALRITY RATIO FOR COMPARING A SCENE OF SCENE 7 AT 0° WITH THE OTHER SCENE IN DATASET USING A* SEARCH TREE ALGORITHM

	0°	10°	20°	30°	40°	50°	60°	70°	80°	90°
Scene 1	21.43	21.43	21.43	21.43	21.43	21.43	25.00	21.43	21.43	21.43
Scene 2	8.33	8.33	8.33	8.33	8.33	8.33	8.33	8.33	8.33	8.33
Scene 3	9.09	9.09	9.09	9.09	9.09	9.09	9.09	9.09	9.09	9.09
Scene 4	50.00	45.45	45.45	45.45	45.45	45.45	45.45	45.45	45.45	45.45
Scene 5	23.08	23.08	23.08	23.08	23.08	23.08	23.08	23.08	23.08	23.08
Scene 6	10.00	10.00	10.00	10.00	10.00	10.00	10.00	10.00	10.00	10.00
Scene 7	100.00	100.00	63.64	63.64	100.00	100.00	100.00	100.00	100.00	100.00
Scene 8	8.33	8.33	8.33	8.33	8.33	11.11	8.33	8.33	8.33	8.33
Scene 9	14.29	14.29	14.29	14.29	14.29	14.29	14.29	14.29	14.29	14.29
Scene 10	14.29	9.09	9.09	9.09	9.09	9.09	9.09	9.09	9.09	9.09

TABLE V. SHOWS THE RESULT OF SIMIALRITY RATIO FOR COMPARING A SCENE OF SCENE 7 AT 0° WITH THE OTHER SCENE IN DATASET USING DECISION TREE WITH LARGEST COMMON SUB-GRAPH ALGORITHM

	0°	10°	20°	30°	40°	50°	60°	70°	80°	90°
Scene 1	28.57	28.57	28.57	28.57	28.57	28.57	33.33	28.57	28.57	28.57
Scene 2	16.67	16.67	16.67	16.67	16.67	16.67	16.67	16.67	16.67	16.67
Scene 3	16.67	16.67	16.67	16.67	16.67	16.67	16.67	16.67	16.67	16.67
Scene 4	50.00	50.00	50.00	50.00	50.00	50.00	50.00	50.00	50.00	50.00
Scene 5	28.57	28.57	28.57	28.57	28.57	28.57	28.57	28.57	28.57	28.57
Scene 6	16.67	16.67	16.67	16.67	16.67	16.67	16.67	16.67	16.67	16.67
Scene 7	100.00	100.00	66.67	66.67	100.00	100.00	100.00	100.00	100.00	100.00
Scene 8	16.67	16.67	16.67	16.67	16.67	20.00	16.67	16.67	16.67	16.67
Scene 9	25.00	25.00	25.00	25.00	25.00	25.00	25.00	25.00	25.00	25.00
Scene 10	25.00	16.67	16.67	16.67	16.67	16.67	16.67	16.67	16.67	16.67

TABLE VI. SHOWS THE RESULT OF SIMIALRITY RATIO FOR COMPARING A SCENE OF SCENE 7 AT 0° WITH THE OTHER SCENE IN DATASET USING PROPOSED ALGORITHM

	0°	10°	20°	30°	40°	50°	60°	70°	80°	90°
Scene 1	21.43	21.43	21.43	21.43	21.43	21.43	25.00	21.43	21.43	21.43
Scene 2	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Scene 3	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Scene 4	40.00	27.27	27.27	27.27	27.27	27.27	27.27	27.27	27.27	27.27
Scene 5	38.46	38.46	38.46	38.46	38.46	38.46	38.46	38.46	38.46	38.46
Scene 6	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Scene 7	100.00	100.00	90.91	90.91	100.00	100.00	100.00	100.00	100.00	100.00
Scene 8	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Scene 9	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Scene 10	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

TABLE VII. SHOWS THE CLASSIFIED MATRIX FOR COMPARING BETWEEN EACH IMAGE WITHIN DATABASE (ONE HUNDRED IMAGES) WHERE THEY ARE CLASSIFIED INTO TEN CLASSES.

	Scene1	Scene2	Scene3	Scene4	Scene5	Scene6	Scene7	Scene8	Scene9	Scene10
Scene1	9	0	0	0	0	0	0	0	0	0
Scene2	0	10	0	0	0	0	0	0	0	0
Scene3	0	0	10	0	0	0	0	0	0	9
Scene4	0	0	0	9	0	0	0	0	0	0
Scene5	0	0	0	0	10	0	0	0	0	0
Scene6	0	0	0	0	0	10	0	0	0	0
Scene7	0	0	0	0	0	0	8	0	0	0
Scene8	0	0	0	0	0	0	0	8	0	0
Scene9	0	0	0	0	0	0	0	0	10	0
Scene10	0	0	9	0	0	0	0	0	0	9

TABLE VII. shows classified matrix for comparing one hundred images with each other and group into ten classes as ten scenes. The scenes three and ten have same object arrangement. The result in classified matrix shows that scene three with ten images and scene ten with ten images have nine images from the same scene that is similar to other images from another scene. This means that there is one image from scene three or scene ten that is not matched with other images of another scene. In case of comparing images in scene one with themselves in other views, the result shows that there is one image not isomorphic with other nine images from the same scene.

6.4. The Experimental Result of Single Plane in Speed

In this experiment, besides the proposed graph matching algorithm, A* and decision tree algorithms from [24] were compared using our databases. Eventhough the accuracy of all three algorithms are not different, the time using for comparing between two graphs was quite different due to the objective in finding common subgraph and graph structure. In paper [24], both A* and decision tree algorithms needed tree model for comparing with input graphs. In their experiment, each video sequence has a tree model. Those models were used in comparing video sequences for finding largest common subgraph. The proposed semantic spatial graph comparison in this dissertation can be used to determine isomorphism and common subgraph between two graphs.

To compare performance of proposed method and the others, each scene with ten images captured from ten different angles were used as a video sequence. A semantic spatial graph of an image from a scene was picked for creating tree model. The video sequences from a scene were compared between using tree model created by A*, decision tree with largest common subgraph detection and our proposed algorithm. The time consuming from comparing between two graphs were shown in TABLE VIII. . The proposed matching algorithm used the least time consuming in comparing between two graphs because the tree model constructing process after retrieving vectorized attributed graph, takes time. Furthermore, the tables of time consuming that was taken in each algorithm from each of experiment were shown in the Appendix. The time in using A* search are the greatest because the decision tree building and the searching method. But DT LCSG is faster than A* because their method in finding largest common sub-graph is faster than A*.

TABLE VIII. SHOWS TIME CONSUMING OF MATCHING BETWEEN TWO GRAPHS(MS).

Algorithm	Mean	Minimum	Maximum	SD
A*	9,440.067	69.031	38,933.39	14,213.46
DT LCSG	9,418.855	69	98933.45	14,545.94
Proposed matching	1.254	0.2713	27.1962	1.5613

For creating decision tree model according to [14], the graph adjacency matrix is used to create permutation of its adjacency matrix to find all possible paths. The time complexities of A* and decision tree algorithms are quite large. The purpose of [14] is to compare video sequences for video indexing. TABLE VIII. TABLE VII. shows the time consuming of three algorithms of matching between two scenes. Decision tree with largest common subgraph detection used less time than A*. However the proposed algorithm used the least time compared with two other methods. The semantic spatial graph structure influence time reduction in matching process. The performance results showed that vectorized attributed graph with proposed matching algorithm can be used in video similarity problem with high speed.

6.5.The Experiment of Single Plane with Real Scenes

In the experiment with real scenes, the dataset from [34] was used. This dataset was collected from the indoor environment with various kinds of rooms. Usually, the furnished rooms such as dining room and living room having a specific kind of pattern. In those rooms, a group of chairs or sofas are placed around the table. Moreover, the office or laboratory has its own pattern of furnishing as well. Figure 22. shows two samples of indoor scenes which are dining room and computer room. For the dining room, the chairs are placed around the table (Number 3 in Figure 22. .a). Its semantic spatial graph is shaped as star. For the computer room, the tables are placed in two parallel rows. Its semantic spatial graph is shaped as line pattern. Each room can be recognized by its semantic spatial graph. Even the objects in the scene are the same, the room can be recognized by the objects' positions. Figure 23. shows a living room scene with two perspective views. The sequence patterns appears in both semantic spatial graphs. Due to the disappearance of object 6 in the second view, the pattern is not exactly the same as first view. Even the sequence is not complete, some sequences can still be recognized.

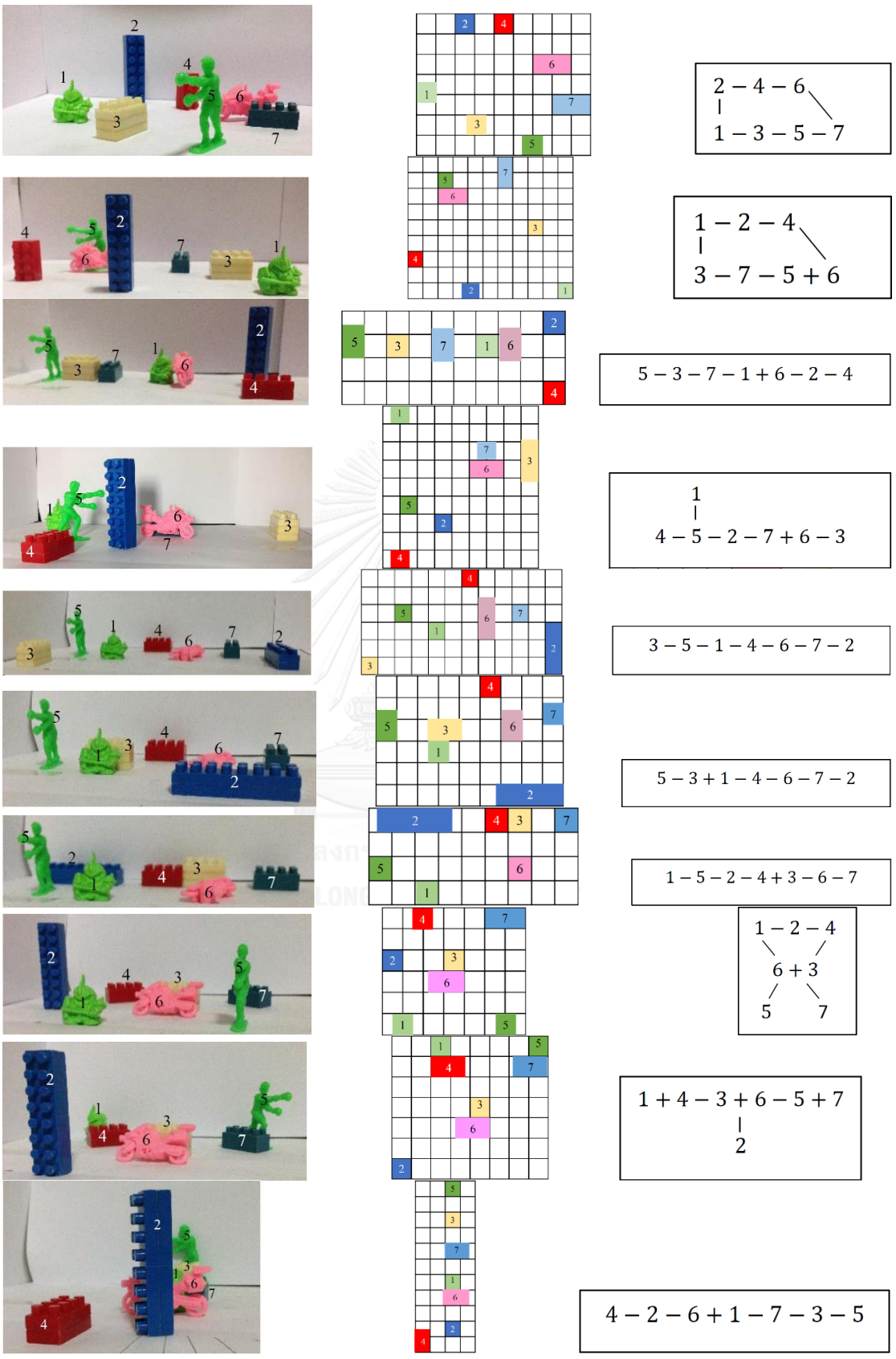
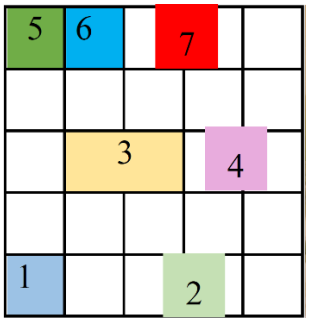


Figure 21. The example of ten scenes in the dataset, their top-view models, and their semantic spatial graphs.

(a)The dining room

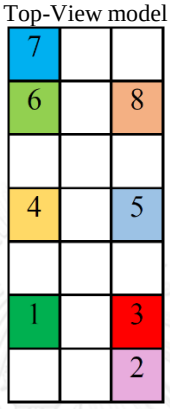


(b) Top-View model

$$\begin{array}{c} 5+6-7 \\ | \\ 1-3-4 \\ | \\ 2 \end{array}$$

(c) A vectorized attributed graph

(d)The computer room



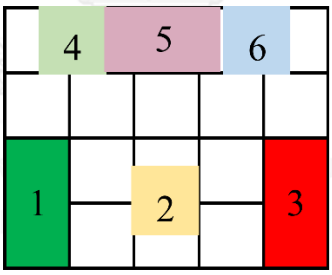
(e) Top-View model

$$2+3-1-4-5-8-7+6$$

(f) A vectorized attributed graph

Figure 22. Examples of rooms (a) a dining room, (d) a computer room, (b) and (e) their top-view model (c) and (f) their vectorized attributed.

(a) The first view

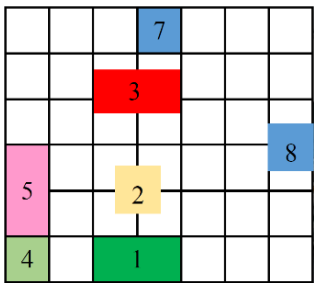


(b) Top-View model

$$\begin{array}{ccc} 4+5+6 \\ | \quad\quad | \\ 1-2-3 \end{array}$$

(c) A vectorized attributed graph

(d) The second view



(e) Top-View model

$$\begin{array}{c} \quad 7 \\ | \\ 4+5-1-2-3-8 \end{array}$$

(f) A vectorized attributed graph

Figure 23. Example of the living room with two views, (a) the first view (b) top-view model of first view (c) vectorized attributed graph of the first view (d) the second view (e) top-view model of first view (f) vectorized attributed graph of the first view.

6.6. The Experiments Results of Multi-plane

The datasets of 10 scenes and 10 frames for each scene in LabelMe databases were selected. From the methodology, the multi-plane images are transformed into vectorized attributed graph also. The example of the vectorized attributed graph which represent the frames in the databases were shown below.

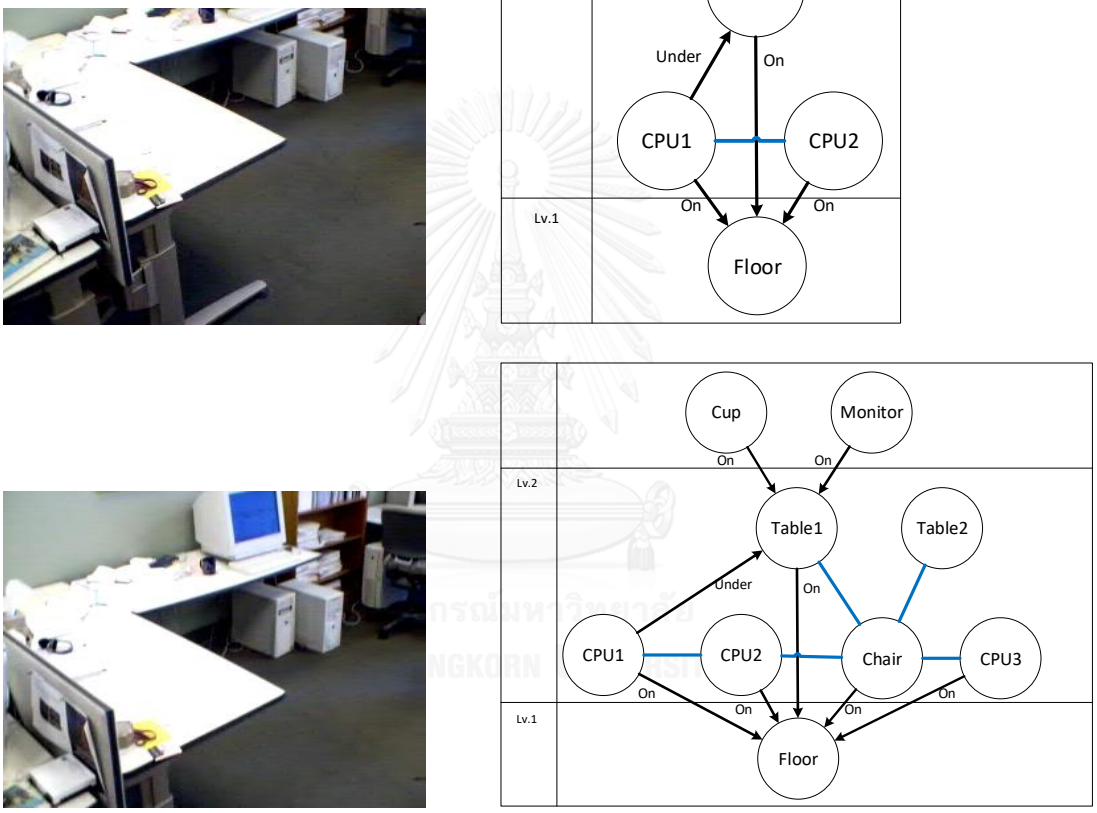


Figure 24. The examples of two frames from a scene in the dataset.

In this work, the objects in the video are labeled and retrieved. To prove the performance of the algorithm, the objects in the frame images are retrieved. In figure 24, the video sequence are captured with swiveling camera. Though there are not any moving objects, the objects that appeared in both frames are slightly different. After retrieving objects, the plane detection will detect planes that are appeared to have objects on them or nothing. A plane is represented as a node and has an attribute that indicates the plane level. Other objects that are on that surface will have their incident edges with attribute on them. The attribute values can be on, under and none where none value indicates the contiguous relationship.

These experiments can be extended to video similarity using the proposed algorithm to compare between frames of two video sequences. If the ratio is quite low, the scene may have been changed rapidly or they are two different scenes. From the different ratio between frames, those videos can be indicated the similarity. To prove the algorithm performance in real dataset the experimental results are shown below. TABLE IX. shows the results of similarity ratio between selected frame 1 and frame 2 and 3 of the same scene in the multi-plane dataset. In the second row, the selected frame 1 of scene 1 were compared with selected frame 1, 2 and 3 of scene 2 respectively. The last row show the ratio of comparing between frame 1 of scene 1 and frame 1, 2 and 3 of scene 3. This table shows the results using A* search with tree model of scene 1 frame 1. In TABLE X. , the results of the same test using decision tree with largest common subgraph and tree model are shown. TABLE XI. shows the result of the same test set using proposed method.

TABLE IX. ACCURACY OF THE GRAPH MATCHING BETWEEN SELECTED FRAME 1 OF SCENE 1 AND OTHER FRAME FROM SCENES 1, 2 AND 3 WITH RATIO USING A* SEARCH.

%	Frame 1	Frame 2	Frame 3
Scene 1	100	100	48
Scene 2	4	3.571429	3.571429
Scene 3	0	0	0

TABLE X. ACCURACY OF THE GRAPH MATCHING BETWEEN SELECTED FRAME 1 OF SCENE 1 AND OTHER FRAME FROM SCENES 1, 2 AND 3 WITH RATIO USING LARGEST COMMON SUB-GRAPH.

%	Frame 1	Frame 2	Frame 3
Scene 1	100	100	48
Scene 2	4	3.571429	3.571429
Scene 3	0	0	0

TABLE XI. ACCURACY OF THE GRAPH MATCHING BETWEEN SELECTED FRAME 1 OF SCENE 1 AND OTHER FRAME FROM SCENES 1, 2 AND 3 WITH RATIO USING PROPOSED METHOD.

%	Frame 1	Frame 2	Frame 3
Scene 1	100	100	48
Scene 2	4	3.571429	3.571429
Scene 3	0	0	0

Though the result of accuracies were the same but the proposed algorithm consumed time less than other methods. Because the structure of graph and the preprocessing that were used in proposed method showed that the graph and the tree model in other method are not necessary. The results for time consuming in millisecond (ms) are shown in tables below. Those time are used in processing the comparison of TABLE IX. , TABLE X. and TABLE XI.

TABLE XII. TIME CONSUMING OF THE GRAPH MATCHING BETWEEN SELECTED FRAME 1 OF SCENE 1 AND OTHER FRAME FROM SCENES 1, 2 AND 3 WITH RATIO USING A* SEARCH.

ms	Frame 1	Frame 2	Frame 3
Scene 1	294.0263	232.0711	51757.42
Scene 2	37931.47	380801.7	378791.8
Scene 3	4114.085	38053.46	38804.14

TABLE XIII. TIME CONSUMING OF THE GRAPH MATCHING BETWEEN SELECTED FRAME 1 OF SCEEN 1 AND OTHER FRAME FROM SCENES 1, 2 AND 3 WITH RATIO USING LARGEST COMMON SUB-GRAPH.

ms	Frame 1	Frame 2	Frame 3
Scene 1	203.6932	157.391	441.9511
Scene 2	5.779347	2.830437	1.930457
Scene 3	1.701723	1.73486	2.025762

TABLE XIV. TIME CONSUMING OF THE GRAPH MATCHING BETWEEN SELECTED FRAME 1 OF SCEEN 1 AND FRAME FROM SCENES 1, 2 AND 3 SCENES WITH RATIO USING PROPOSED METHOD.

ms	Frame 1	Frame 2	Frame 3
Scene 1	16.45972	5.11086	3.761182
Scene 2	2.627043	1.276432	1.247974
Scene 3	0.690468	0.6942	0.677871

7. CONCLUSION

In the top-view model algorithm, the algorithm starts with placing each object into grid unit space. In the best and worst cases, the algorithm takes $O(n^2)$ where n is the number of objects in the top-view model. In the check value algorithm, all objects are checked. For the worst case this will take $O(n)$, but, for best case, this method takes $O(1)$ for time consuming. Thus, for the worst case of top-view model algorithm, the complexity of the best case is $O(n^3)$. While, the complexity is $O(n^2)$. In conclusion, the top-view model complexity is less than $O(n^3)$. The time complexity for finding semantic spatial graph is $O(n)$ for finding two or three nearest objects. In graph comparison, both adjacency matrices of two graphs are checked through all objects. This cost the time complexity to $O(n^2)$. In summary, the worst case of proposed algorithm is $O(n^3)$. The best case is $O(n^2)$.

The top-view modeling algorithm to project scene in videos and images on the simulated space is presented. The simulated space provide the valuable information for producing special features such as graph. Graph with special structure is used to show the existence of objects and their relationships in a scene. With the interval-based spatial logic, the objects' positions are encoded into two alignment parameters using spatial logic function that demonstrates both horizontal and vertical alignments between two objects. Later the top-view model of that scene is decoded from those parameters. The same scene will produce the same top-view model though the images or videos that can be viewed in different angles. Furthermore, top-view model plays a significant role in expressing the semantic spatial graph. In other words, the same scene can be identified from the existence of semantic spatial graph. The examples show that the definitions and theories are practically usable. This also proves that the nearest objects relation is considerable as a recognizable pattern.

Besides, there are possibilities to enhance the flexibility in measuring the similarity between scenes. The exceptions sometimes can be accepted using partial graph matching in the case that partial scene is more interested than the whole scene. In this dissertation, the research emphasized the representation of scenes in videos and images. If the temporal topology is concerned and defined by functions and extended vectorized attributted graph, the better results should be obtained.

APPENDIX

The HSL space color chart from Wikipedia site [35].

	***H* = 180°** (Cyan)					***H* = 0°** (Red)			
L \ *S*	1	¾	½	¼	0	¼	½	¾	1
1									
⅞									
¾									
⅝									
½									
⅜									
¼									
⅛									
0									

	***H* = 210°** (Blue-Cyan)					***H* = 30°** (Yellow-Red)			
L \ *S*	1	¾	½	¼	0	¼	½	¾	1
1									
⅞									
¾									
⅝									
½									
⅜									
¼									
⅛									
0									

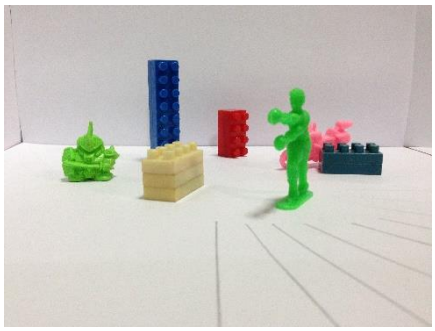
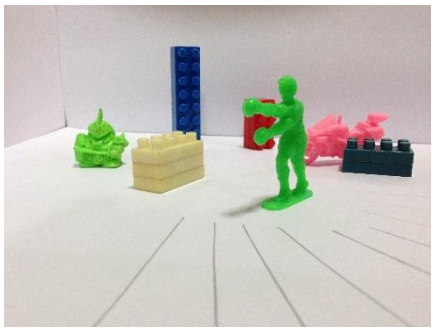
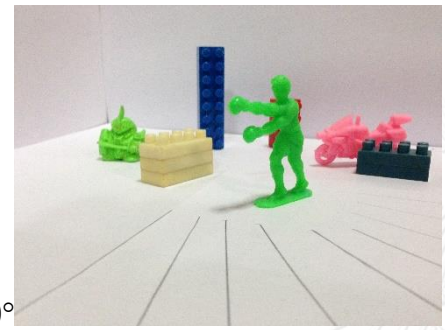
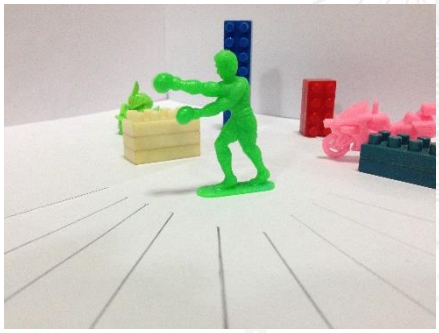
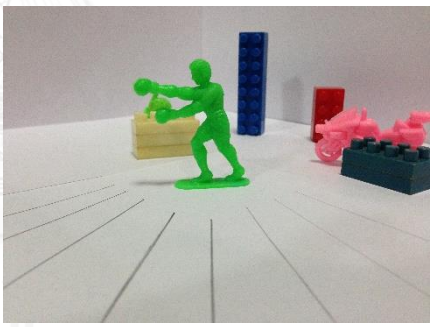
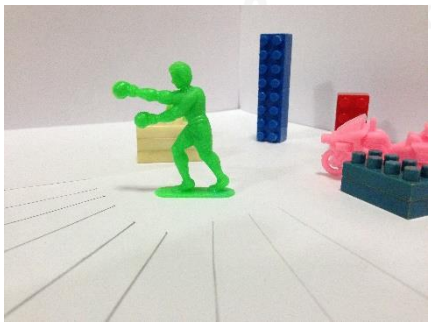
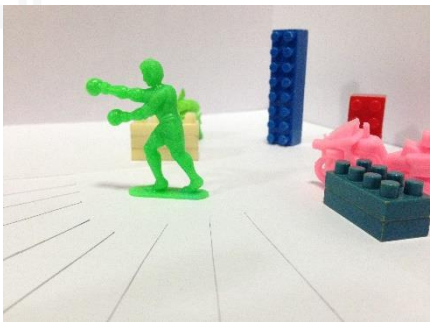
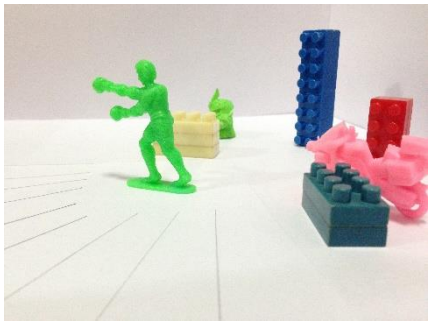
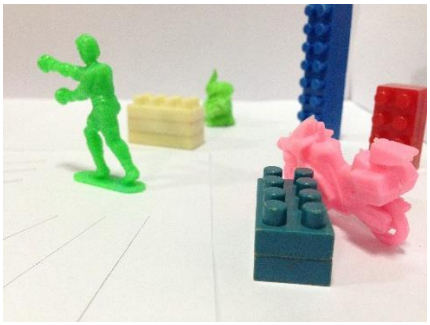
	***H* = 240°** (Blue)					***H* = 60°** (Yellow)			
L \ *S*	1	¾	½	¼	0	¼	½	¾	1
1									
⅞									
¾									
⅝									
½									
⅜									
¼									
⅛									
0									

	***H* = 330°** (Red-Magenta)					***H* = 150°** (Cyan-Green)			
L \ *S*	1	¾	½	¼	0	¼	½	¾	1
1									
⅞									
¾									
⅝									
½									
⅜									
¼									
⅛									
0									

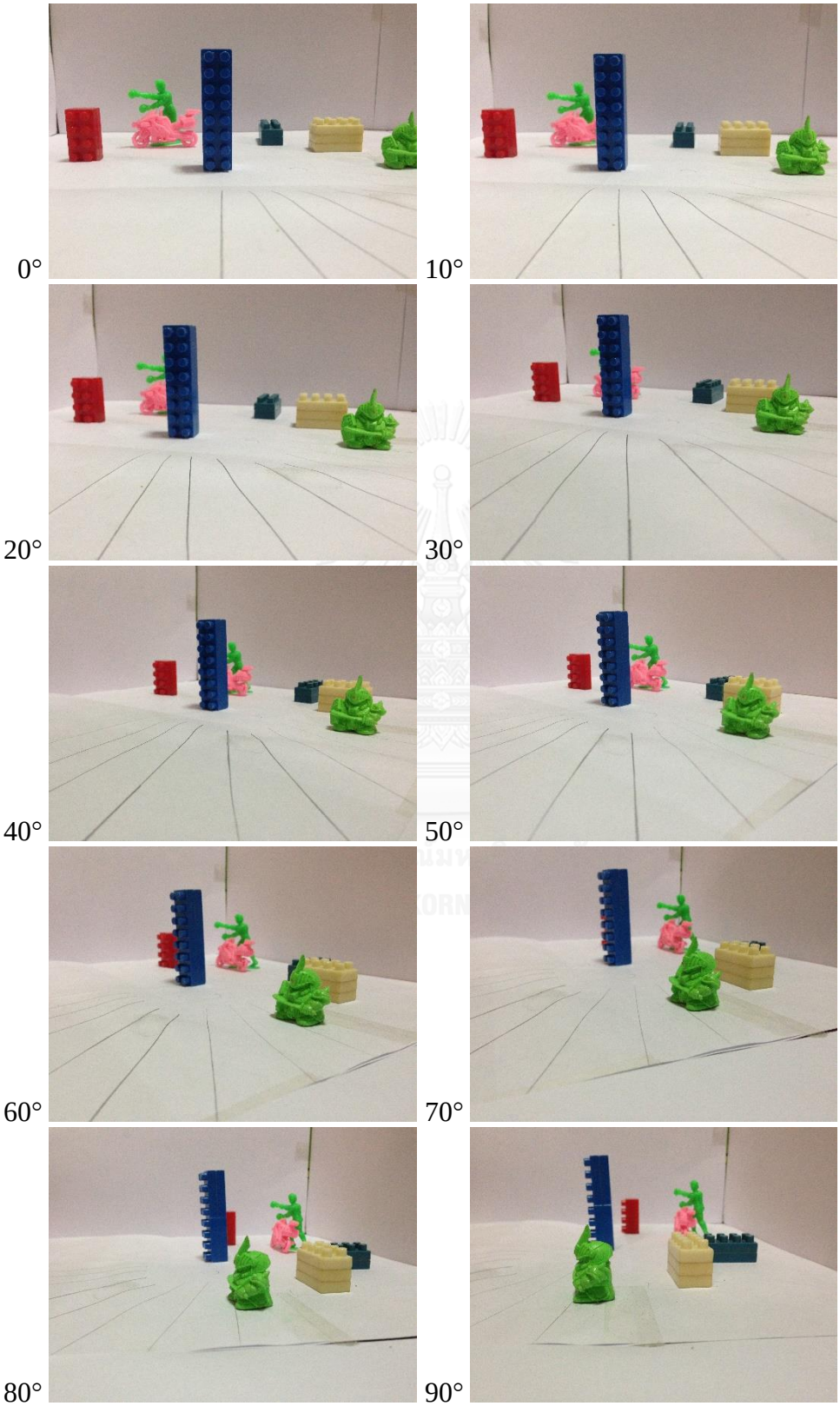
	***H* = 300°** (Magenta)					***H* = 120°** (Green)			
L \ *S*	1	¾	½	¼	0	¼	½	¾	1
1									
⅞									
¾									
⅝									
½									
⅜									
¼									
⅛									
0									

	***H* = 270°** (Magenta-Blue)					***H* = 90°** (Green-Yellow)			
L \ *S*	1	¾	½	¼	0	¼	½	¾	1
1									
⅞									
¾									
⅝									
½									
⅜									
¼									
⅛									
0									

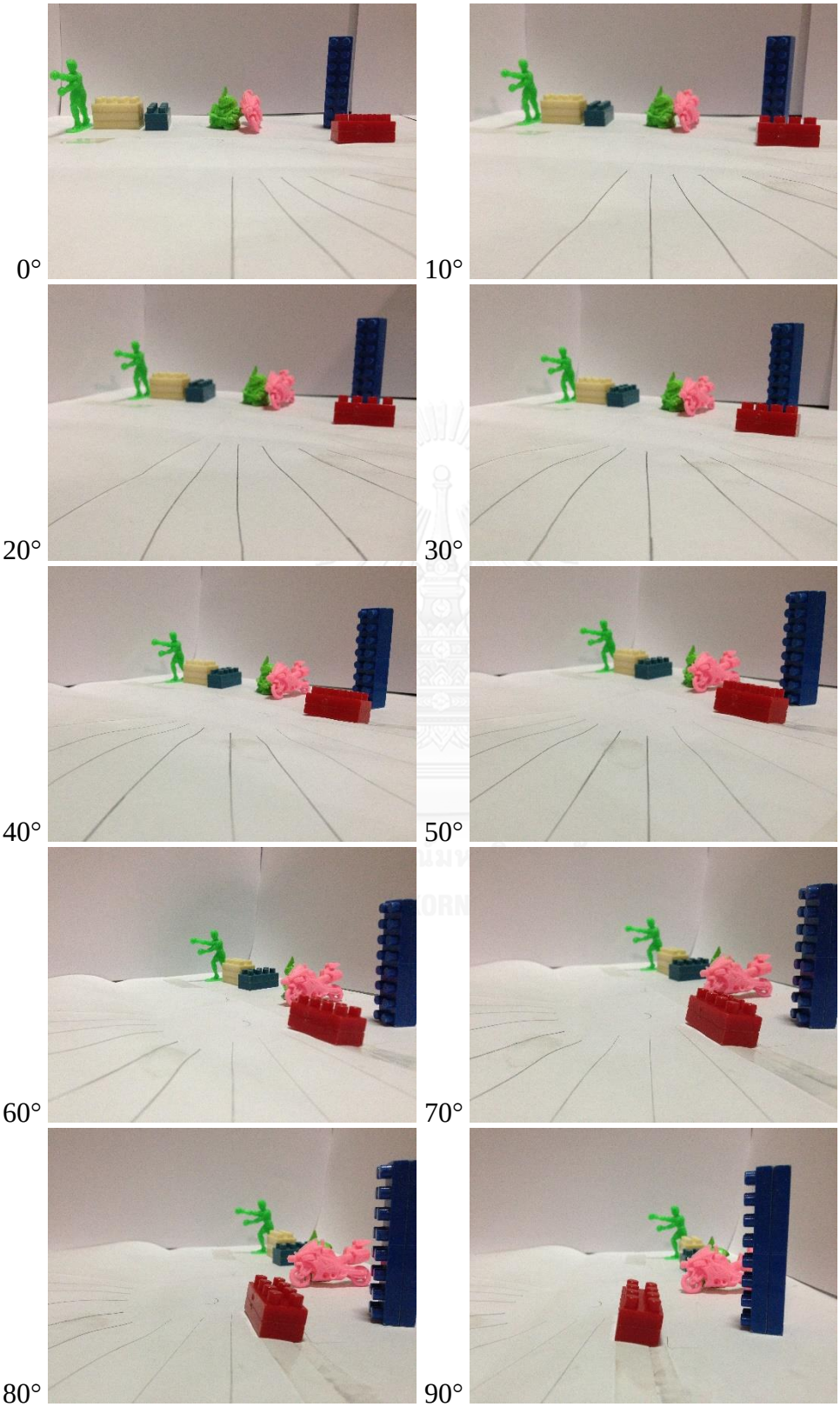
The frames in video sequences as the dataset for single plane experiment are shown below where each frame are captured from different angles with 10° differences.
Scene 1.

0° 
10° 
20° 
30° 
40° 
50° 
60° 
70° 
80° 
90° 

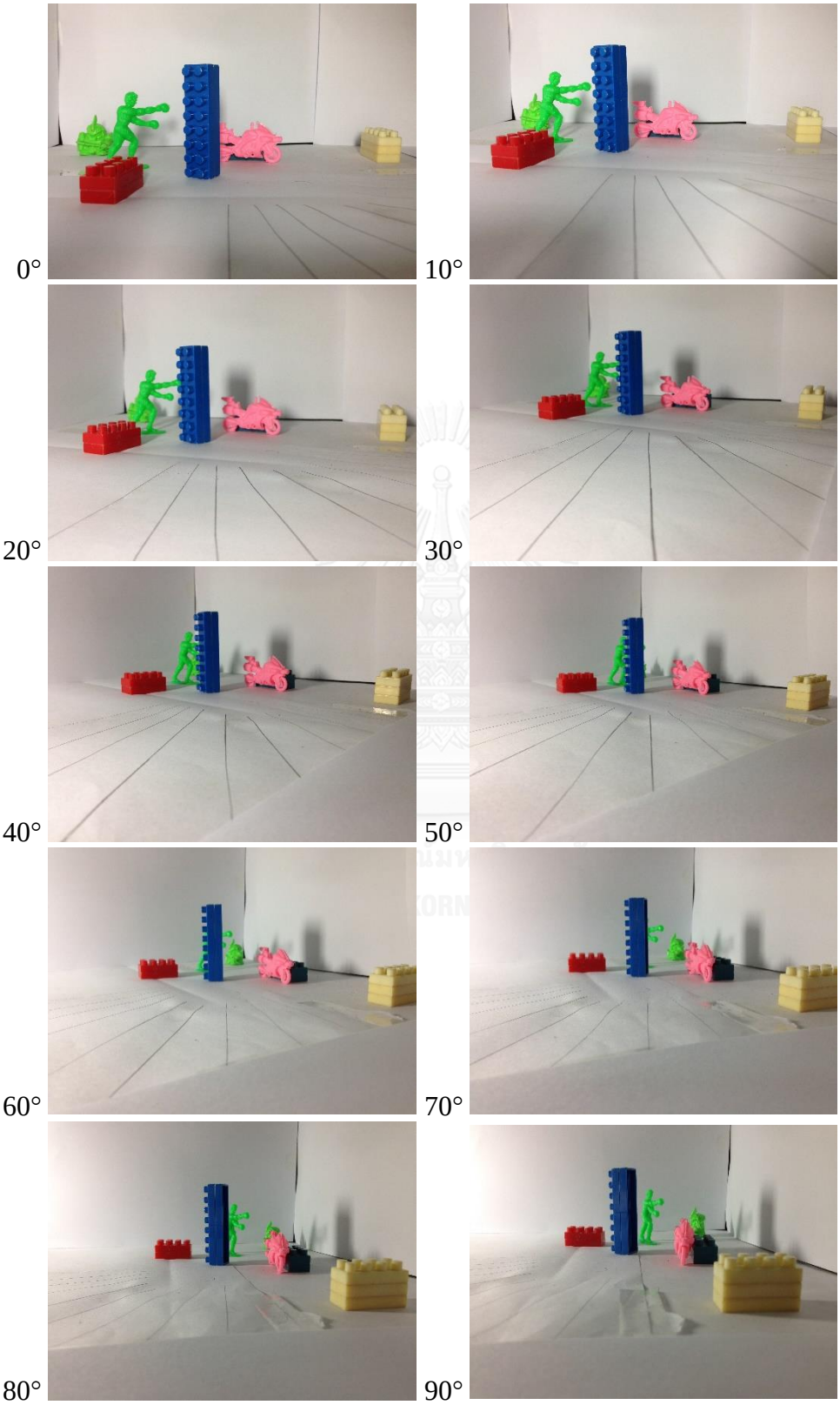
Scene 2.



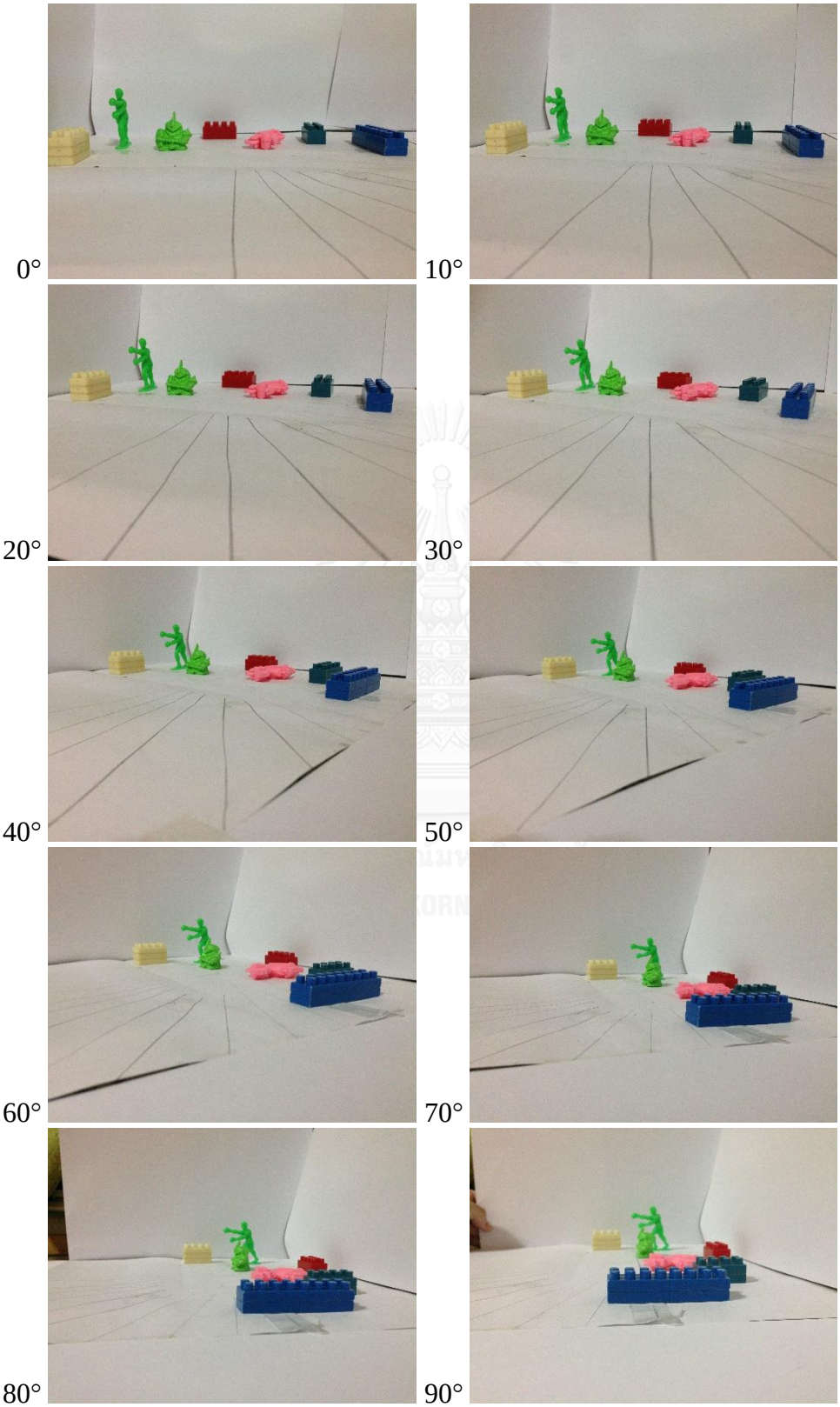
Scene 3.



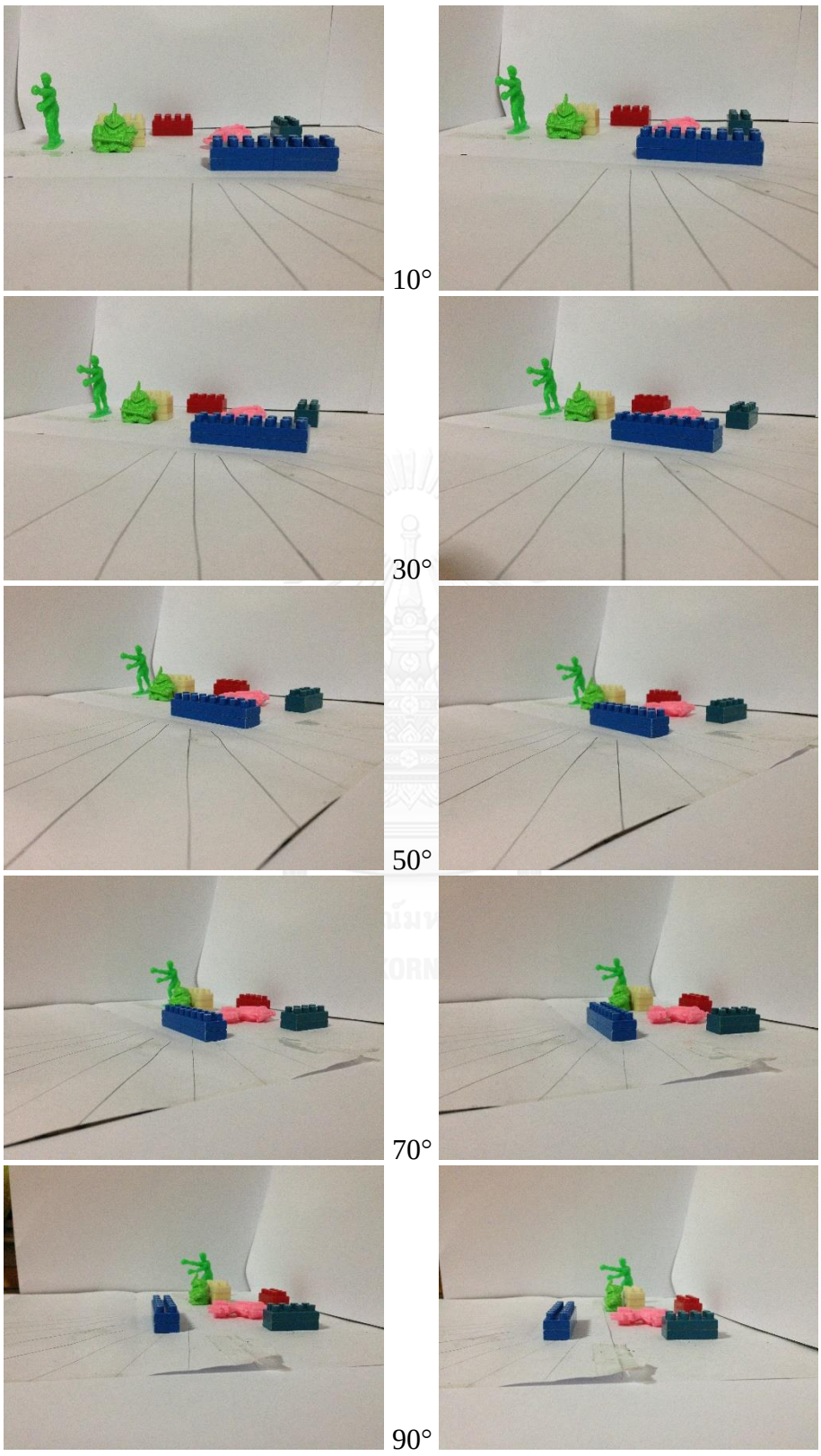
Scene 4.



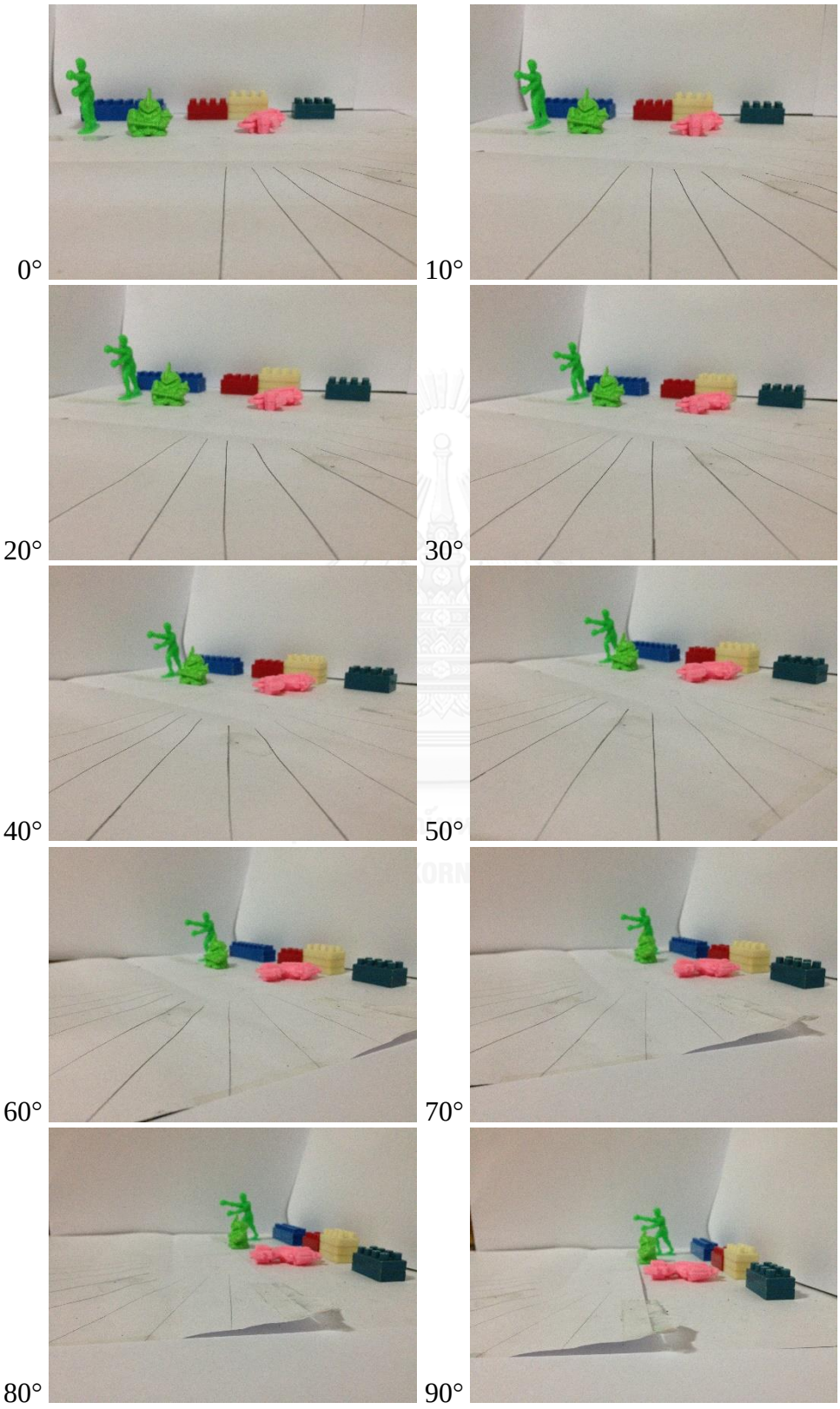
Scene 5.



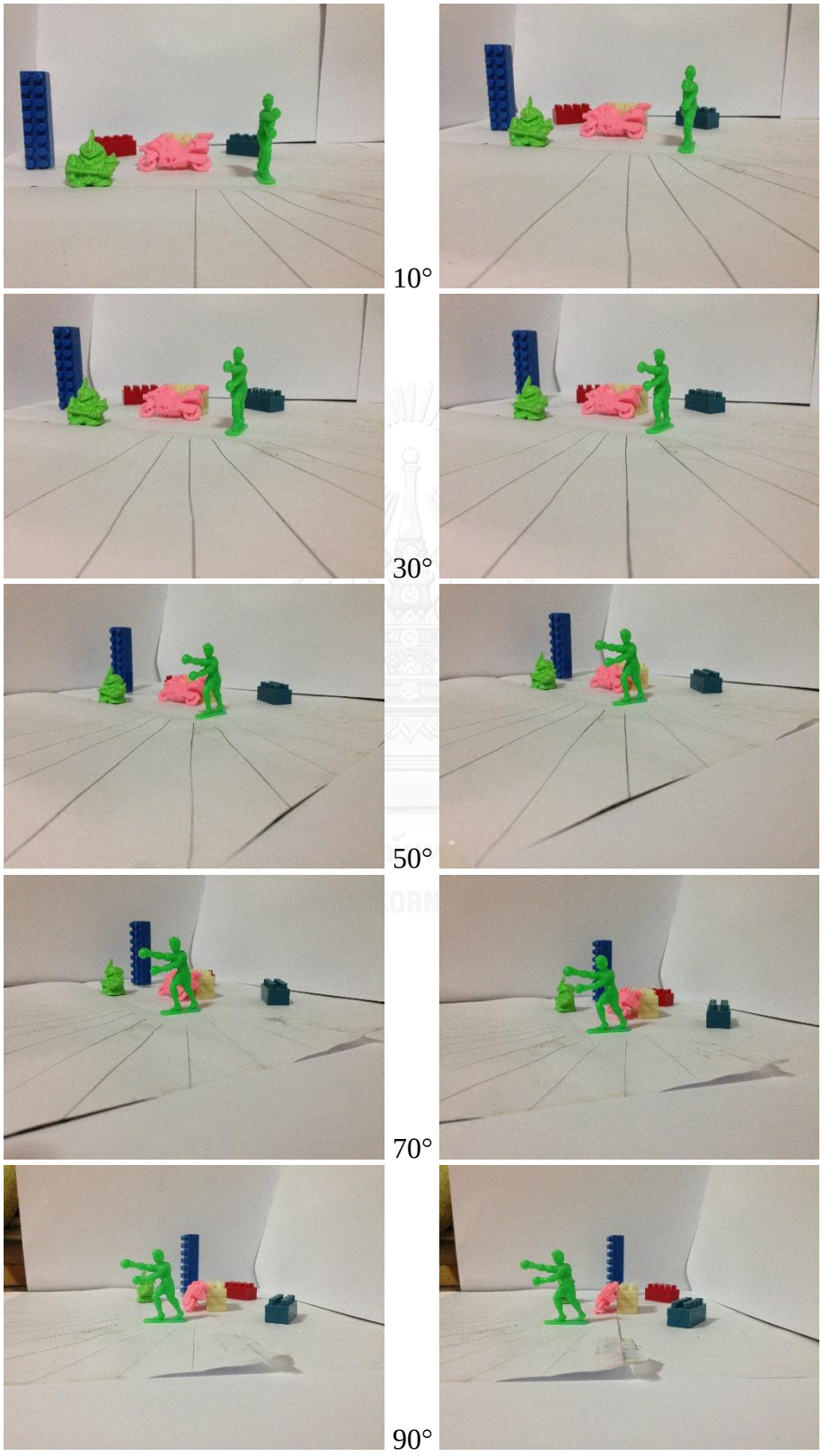
Scene 6.



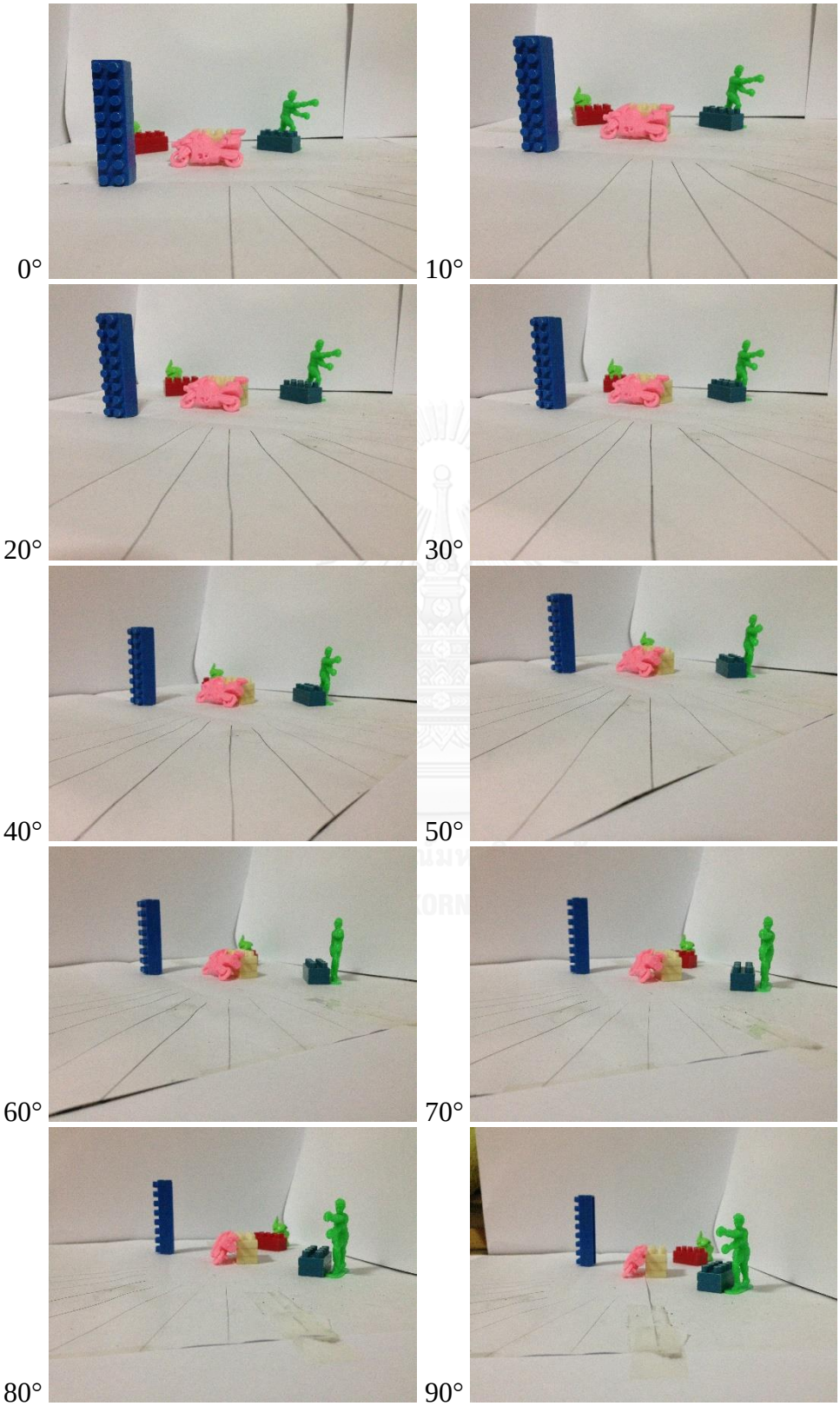
Scene 7.



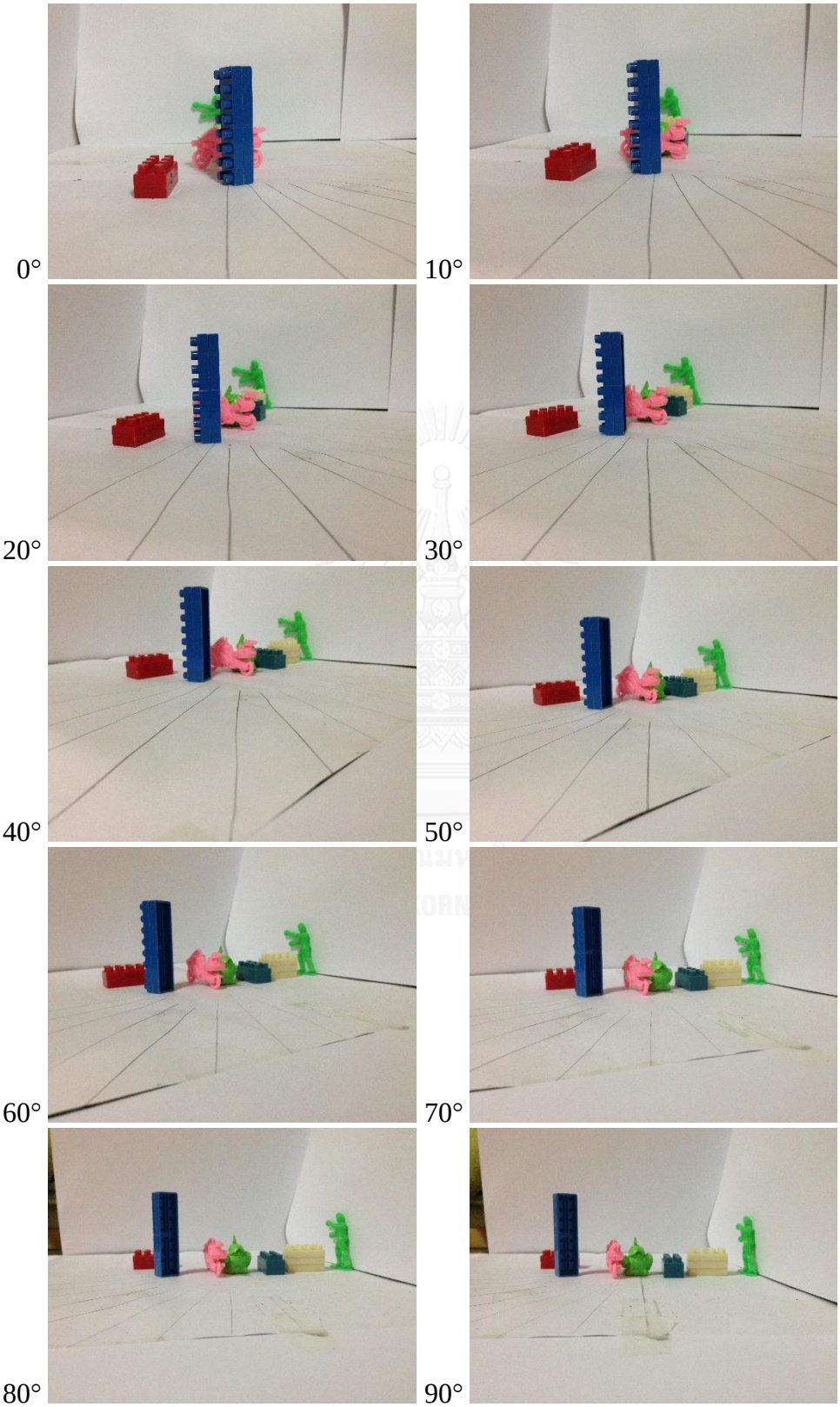
Scene 8.



Scene 9.



Scene 10.



The frames in video sequences as the dataset for multi-level plane experiments are shown below which each sequence containing 10 frames.

Scene1.

Frame 1 

Frame 2 

Frame 3 

Frame 4 

Frame 5 

Frame 6 

Frame 7 

Frame 8 

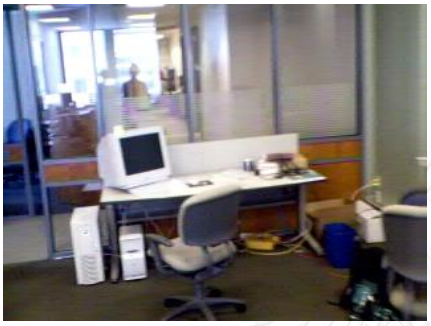
Frame 9 

Frame 10 

Scene 2.

Frame 1 

Frame 2 

Frame 3 

Frame 4 

Frame 5 

Frame 6 

Frame 7 

Frame 8 

Frame 9 

Frame 10 

Scene 3.

Frame 1 

Frame 2 

Frame 3 

Frame 4 

Frame 5 

Frame 6 

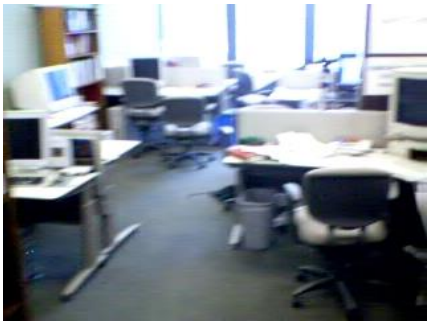
Frame 7 

Frame 8 

Frame 9 

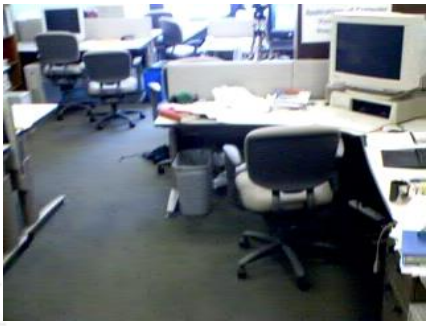
Frame 10 

Scene 4.

Frame 1 

Frame 2 

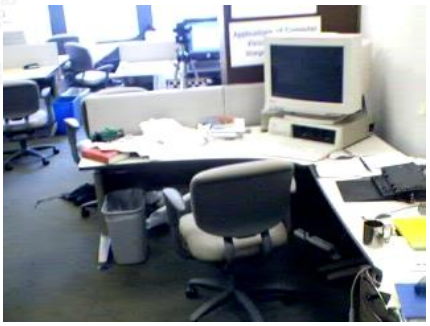
Frame 3 

Frame 4 

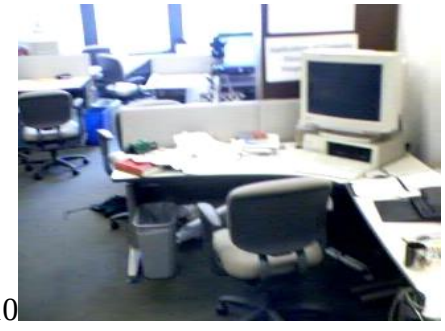
Frame 5 

Frame 6 

Frame 7 

Frame 8 

Frame 9 

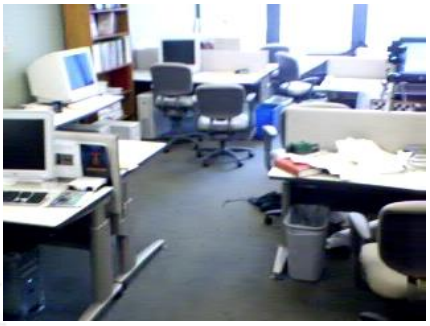
Frame 10 

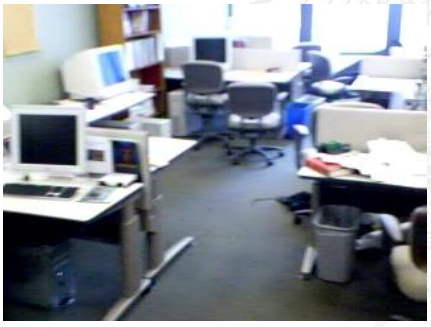
Scene 5.

Frame 1 

Frame 2 

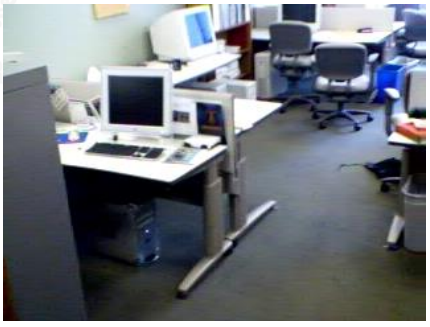
Frame 3 

Frame 4 

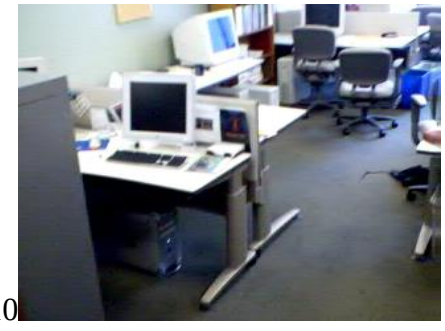
Frame 5 

Frame 6 

Frame 7 

Frame 8 

Frame 9 

Frame 10 

Scene 6.

Frame 1 

Frame 2 

Frame 3 

Frame 4 

Frame 5 

Frame 6 

Frame 7 

Frame 8 

Frame 9 

Frame 10 

Scene 7.

Frame 1 

Frame 2 

Frame 3 

Frame 4 

Frame 5 

Frame 6 

Frame 7 

Frame 8 

Frame 9 

Frame 10 

Scene 8.

Frame 1 

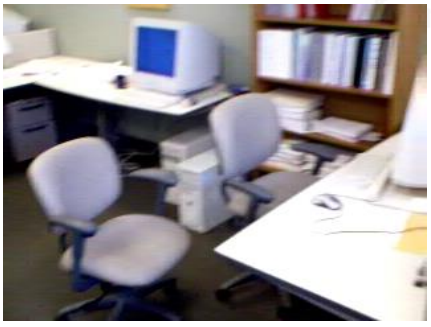
Frame 2 

Frame 3 

Frame 4 

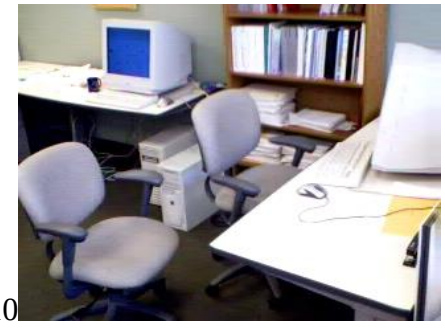
Frame 5 

Frame 6 

Frame 7 

Frame 8 

Frame 9 

Frame 10 

Scene 9.

Frame 1 

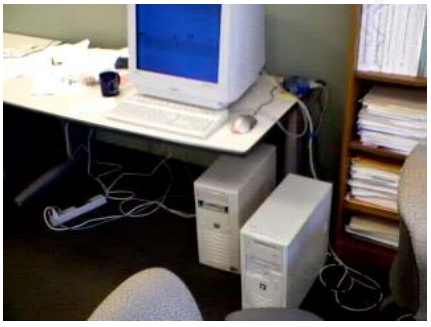
Frame 2 

Frame 3 

Frame 4 

Frame 5 

Frame 6 

Frame 7 

Frame 8 

Frame 9 

Frame 10 

Scene 10.

Frame 1 

Frame 2 

Frame 3 

Frame 4 

Frame 5 

Frame 6 

Frame 7 

Frame 8 

Frame 9 

Frame 10 

Result of single plane dataset in terms of accuracy

1.) The results of similarity ratio between scene representative image at 0° and whole dataset using A* search with tree model
Image model of Scene 1 as compared image

	0°	10°	20°	30°	40°	50°	60°	70°	80°	90°
Scene 1	100.00	100.00	100.00	100.00	100.00	100.00	91.67	100.00	100.00	100.00
Scene 2	58.33	58.33	58.33	58.33	58.33	58.33	58.33	58.33	58.33	58.33
Scene 3	27.27	27.27	27.27	27.27	27.27	27.27	27.27	27.27	27.27	27.27
Scene 4	30.00	9.09	9.09	9.09	9.09	9.09	9.09	9.09	9.09	9.09
Scene 5	38.46	38.46	38.46	38.46	38.46	38.46	38.46	38.46	38.46	38.46
Scene 6	30.00	30.00	30.00	30.00	30.00	30.00	30.00	30.00	30.00	30.00
Scene 7	27.27	27.27	27.27	27.27	27.27	27.27	27.27	27.27	27.27	27.27
Scene 8	41.67	41.67	41.67	41.67	41.67	33.33	41.67	41.67	41.67	41.67
Scene 9	14.29	14.29	14.29	14.29	14.29	14.29	14.29	14.29	14.29	14.29
Scene 10	42.86	27.27	27.27	27.27	27.27	27.27	27.27	27.27	27.27	27.27

Image model of Scene 2 as compared image

	0°	10°	20°	30°	40°	50°	60°	70°	80°	90°
Scene 1	50.00	50.00	50.00	50.00	50.00	50.00	25.00	50.00	50.00	50.00
Scene 2	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
Scene 3	27.27	27.27	27.27	27.27	27.27	27.27	27.27	27.27	27.27	27.27
Scene 4	30.00	9.09	9.09	9.09	9.09	9.09	9.09	9.09	9.09	9.09
Scene 5	7.69	7.69	7.69	7.69	7.69	7.69	7.69	7.69	7.69	7.69
Scene 6	10.00	10.00	10.00	10.00	10.00	10.00	10.00	10.00	10.00	10.00
Scene 7	9.09	9.09	9.09	9.09	9.09	9.09	9.09	9.09	9.09	9.09
Scene 8	41.67	41.67	41.67	41.67	41.67	33.33	41.67	41.67	41.67	41.67
Scene 9	14.29	14.29	14.29	14.29	14.29	14.29	14.29	14.29	14.29	14.29
Scene 10	42.86	27.27	27.27	27.27	27.27	27.27	27.27	27.27	27.27	27.27

Image model of Scene 3 as compared image

	0°	10°	20°	30°	40°	50°	60°	70°	80°	90°
Scene 1	21.43	21.43	21.43	21.43	21.43	21.43	25.00	21.43	21.43	21.43
Scene 2	25.00	25.00	25.00	25.00	25.00	25.00	25.00	25.00	25.00	25.00
Scene 3	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
Scene 4	30.00	9.09	9.09	9.09	9.09	9.09	9.09	9.09	9.09	9.09
Scene 5	23.08	23.08	23.08	23.08	23.08	23.08	23.08	23.08	23.08	23.08
Scene 6	10.00	10.00	10.00	10.00	10.00	10.00	10.00	10.00	10.00	10.00
Scene 7	9.09	9.09	9.09	9.09	9.09	9.09	9.09	9.09	9.09	9.09
Scene 8	25.00	25.00	25.00	25.00	25.00	11.11	25.00	25.00	25.00	25.00
Scene 9	14.29	14.29	14.29	14.29	14.29	14.29	14.29	14.29	14.29	14.29
Scene 10	42.86	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00

Image model of Scene 4 as compared image

	0°	10°	20°	30°	40°	50°	60°	70°	80°	90°
Scene 1	7.14	7.14	7.14	7.14	7.14	7.14	8.33	7.14	7.14	7.14
Scene 2	8.33	8.33	8.33	8.33	8.33	8.33	8.33	8.33	8.33	8.33
Scene 3	9.09	9.09	9.09	9.09	9.09	9.09	9.09	9.09	9.09	9.09
Scene 4	50.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
Scene 5	23.08	23.08	23.08	23.08	23.08	23.08	23.08	23.08	23.08	23.08
Scene 6	10.00	10.00	10.00	10.00	10.00	10.00	10.00	10.00	10.00	10.00
Scene 7	45.45	45.45	45.45	45.45	45.45	45.45	45.45	45.45	45.45	45.45
Scene 8	8.33	8.33	8.33	8.33	8.33	11.11	8.33	8.33	8.33	8.33
Scene 9	14.29	14.29	14.29	14.29	14.29	14.29	14.29	14.29	14.29	14.29
Scene 10	14.29	9.09	9.09	9.09	9.09	9.09	9.09	9.09	9.09	9.09

Image model of Scene 5 as compared image

	0°	10°	20°	30°	40°	50°	60°	70°	80°	90°
Scene 1	35.71	35.71	35.71	35.71	35.71	35.71	41.67	35.71	35.71	35.71
Scene 2	8.33	8.33	8.33	8.33	8.33	8.33	8.33	8.33	8.33	8.33
Scene 3	27.27	27.27	27.27	27.27	27.27	27.27	27.27	27.27	27.27	27.27
Scene 4	30.00	27.27	27.27	27.27	27.27	27.27	27.27	27.27	27.27	27.27
Scene 5	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
Scene 6	30.00	30.00	30.00	30.00	30.00	30.00	30.00	30.00	30.00	30.00
Scene 7	27.27	27.27	27.27	27.27	27.27	27.27	27.27	27.27	27.27	27.27
Scene 8	8.33	8.33	8.33	8.33	8.33	11.11	8.33	8.33	8.33	8.33
Scene 9	14.29	14.29	14.29	14.29	14.29	14.29	14.29	14.29	14.29	14.29
Scene 10	14.29	27.27	27.27	27.27	27.27	27.27	27.27	27.27	27.27	27.27

Image model of Scene 6 as compared image

	0°	10°	20°	30°	40°	50°	60°	70°	80°	90°
Scene 1	21.43	21.43	21.43	21.43	21.43	21.43	25.00	21.43	21.43	21.43
Scene 2	8.33	8.33	8.33	8.33	8.33	8.33	8.33	8.33	8.33	8.33
Scene 3	9.09	9.09	9.09	9.09	9.09	9.09	9.09	9.09	9.09	9.09
Scene 4	10.00	9.09	9.09	9.09	9.09	9.09	9.09	9.09	9.09	9.09
Scene 5	23.08	23.08	23.08	23.08	23.08	23.08	23.08	23.08	23.08	23.08
Scene 6	70.00	70.00	70.00	70.00	70.00	70.00	70.00	70.00	70.00	70.00
Scene 7	9.09	9.09	9.09	9.09	9.09	9.09	9.09	9.09	9.09	9.09
Scene 8	8.33	8.33	8.33	8.33	8.33	11.11	8.33	8.33	8.33	8.33
Scene 9	14.29	14.29	14.29	14.29	14.29	14.29	14.29	14.29	14.29	14.29
Scene 10	14.29	9.09	9.09	9.09	9.09	9.09	9.09	9.09	9.09	9.09

Image model of Scene 7 as compared image

	0°	10°	20°	30°	40°	50°	60°	70°	80°	90°
Scene 1	21.43	21.43	21.43	21.43	21.43	21.43	25.00	21.43	21.43	21.43
Scene 2	8.33	8.33	8.33	8.33	8.33	8.33	8.33	8.33	8.33	8.33
Scene 3	9.09	9.09	9.09	9.09	9.09	9.09	9.09	9.09	9.09	9.09
Scene 4	50.00	45.45	45.45	45.45	45.45	45.45	45.45	45.45	45.45	45.45
Scene 5	23.08	23.08	23.08	23.08	23.08	23.08	23.08	23.08	23.08	23.08
Scene 6	10.00	10.00	10.00	10.00	10.00	10.00	10.00	10.00	10.00	10.00
Scene 7	100.00	100.00	63.64	63.64	100.00	100.00	100.00	100.00	100.00	100.00
Scene 8	8.33	8.33	8.33	8.33	8.33	11.11	8.33	8.33	8.33	8.33
Scene 9	14.29	14.29	14.29	14.29	14.29	14.29	14.29	14.29	14.29	14.29
Scene 10	14.29	9.09	9.09	9.09	9.09	9.09	9.09	9.09	9.09	9.09

Image model of Scene 8 as compared image

	0°	10°	20°	30°	40°	50°	60°	70°	80°	90°
Scene 1	35.71	35.71	35.71	35.71	35.71	35.71	25.00	35.71	35.71	35.71
Scene 2	41.67	41.67	41.67	41.67	41.67	41.67	41.67	41.67	41.67	41.67
Scene 3	27.27	27.27	27.27	27.27	27.27	27.27	27.27	27.27	27.27	27.27
Scene 4	30.00	9.09	9.09	9.09	9.09	9.09	9.09	9.09	9.09	9.09
Scene 5	7.69	7.69	7.69	7.69	7.69	7.69	7.69	7.69	7.69	7.69
Scene 6	10.00	10.00	10.00	10.00	10.00	10.00	10.00	10.00	10.00	10.00
Scene 7	9.09	9.09	9.09	9.09	9.09	9.09	9.09	9.09	9.09	9.09
Scene 8	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
Scene 9	14.29	14.29	14.29	14.29	14.29	14.29	14.29	14.29	14.29	14.29
Scene 10	42.86	27.27	27.27	27.27	27.27	27.27	27.27	27.27	27.27	27.27

Image model of Scene 9 as compared image

	0°	10°	20°	30°	40°	50°	60°	70°	80°	90°
Scene 1	7.14	7.14	7.14	7.14	7.14	7.14	8.33	7.14	7.14	7.14
Scene 2	8.33	8.33	8.33	8.33	8.33	8.33	8.33	8.33	8.33	8.33
Scene 3	9.09	9.09	9.09	9.09	9.09	9.09	9.09	9.09	9.09	9.09
Scene 4	10.00	9.09	9.09	9.09	9.09	9.09	9.09	9.09	9.09	9.09
Scene 5	7.69	7.69	7.69	7.69	7.69	7.69	7.69	7.69	7.69	7.69
Scene 6	10.00	10.00	10.00	10.00	10.00	10.00	10.00	10.00	10.00	10.00
Scene 7	9.09	9.09	9.09	9.09	9.09	9.09	9.09	9.09	9.09	9.09
Scene 8	8.33	8.33	8.33	8.33	8.33	11.11	8.33	8.33	8.33	8.33
Scene 9	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
Scene 10	14.29	9.09	9.09	9.09	9.09	9.09	9.09	9.09	9.09	9.09

Image model of Scene 10 as compared image

	0°	10°	20°	30°	40°	50°	60°	70°	80°	90°
Scene 1	21.43	21.43	21.43	21.43	21.43	21.43	25.00	21.43	21.43	21.43
Scene 2	25.00	25.00	25.00	25.00	25.00	25.00	25.00	25.00	25.00	25.00
Scene 3	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
Scene 4	30.00	9.09	9.09	9.09	9.09	9.09	9.09	9.09	9.09	9.09
Scene 5	23.08	23.08	23.08	23.08	23.08	23.08	23.08	23.08	23.08	23.08
Scene 6	10.00	10.00	10.00	10.00	10.00	10.00	10.00	10.00	10.00	10.00
Scene 7	9.09	9.09	9.09	9.09	9.09	9.09	9.09	9.09	9.09	9.09
Scene 8	25.00	25.00	25.00	25.00	25.00	11.11	25.00	25.00	25.00	25.00
Scene 9	14.29	14.29	14.29	14.29	14.29	14.29	14.29	14.29	14.29	14.29
Scene 10	42.86	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00

2.) The results of similarity ratio between scene representative image at 0° and whole dataset using Decision tree Largest Common Sub-graph with tree model

Image model of Scene 1 as compared image

	0°	10°	20°	30°	40°	50°	60°	70°	80°	90°
Scene 1	100.00	100.00	100.00	100.00	100.00	100.00	83.33	100.00	100.00	100.00
Scene 2	66.67	66.67	66.67	66.67	66.67	66.67	66.67	66.67	66.67	66.67
Scene 3	33.33	33.33	33.33	33.33	33.33	33.33	33.33	33.33	33.33	33.33
Scene 4	33.33	16.67	16.67	16.67	16.67	16.67	16.67	16.67	16.67	16.67
Scene 5	42.86	42.86	42.86	42.86	42.86	42.86	42.86	42.86	42.86	42.86
Scene 6	33.33	33.33	33.33	33.33	33.33	33.33	33.33	33.33	33.33	33.33
Scene 7	33.33	33.33	33.33	33.33	33.33	33.33	33.33	33.33	33.33	33.33
Scene 8	50.00	50.00	50.00	50.00	50.00	40.00	50.00	50.00	50.00	50.00
Scene 9	25.00	25.00	25.00	25.00	25.00	25.00	25.00	25.00	25.00	25.00
Scene 10	50.00	33.33	33.33	33.33	33.33	33.33	33.33	33.33	33.33	33.33

Image model of Scene 2 as compared image

	0°	10°	20°	30°	40°	50°	60°	70°	80°	90°
Scene 1	57.14	57.14	57.14	57.14	57.14	57.14	33.33	57.14	57.14	57.14
Scene 2	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
Scene 3	33.33	33.33	33.33	33.33	33.33	33.33	33.33	33.33	33.33	33.33
Scene 4	33.33	16.67	16.67	16.67	16.67	16.67	16.67	16.67	16.67	16.67
Scene 5	14.29	14.29	14.29	14.29	14.29	14.29	14.29	14.29	14.29	14.29
Scene 6	16.67	16.67	16.67	16.67	16.67	16.67	16.67	16.67	16.67	16.67
Scene 7	16.67	16.67	16.67	16.67	16.67	16.67	16.67	16.67	16.67	16.67
Scene 8	50.00	50.00	50.00	50.00	50.00	40.00	50.00	50.00	50.00	50.00
Scene 9	25.00	25.00	25.00	25.00	25.00	25.00	25.00	25.00	25.00	25.00
Scene 10	50.00	33.33	33.33	33.33	33.33	33.33	33.33	33.33	33.33	33.33

Image model of Scene 3 as compared image

	0°	10°	20°	30°	40°	50°	60°	70°	80°	90°
Scene 1	28.57	28.57	28.57	28.57	28.57	28.57	33.33	28.57	28.57	28.57
Scene 2	33.33	33.33	33.33	33.33	33.33	33.33	33.33	33.33	33.33	33.33
Scene 3	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
Scene 4	33.33	16.67	16.67	16.67	16.67	16.67	16.67	16.67	16.67	16.67
Scene 5	28.57	28.57	28.57	28.57	28.57	28.57	28.57	28.57	28.57	28.57
Scene 6	16.67	16.67	16.67	16.67	16.67	16.67	16.67	16.67	16.67	16.67
Scene 7	16.67	16.67	16.67	16.67	16.67	16.67	16.67	16.67	16.67	16.67
Scene 8	33.33	33.33	33.33	33.33	33.33	20.00	33.33	33.33	33.33	33.33
Scene 9	25.00	25.00	25.00	25.00	25.00	25.00	25.00	25.00	25.00	25.00
Scene 10	50.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00

Image model of Scene 4 as compared image

	0°	10°	20°	30°	40°	50°	60°	70°	80°	90°
Scene 1	14.29	14.29	14.29	14.29	14.29	14.29	16.67	14.29	14.29	14.29
Scene 2	16.67	16.67	16.67	16.67	16.67	16.67	16.67	16.67	16.67	16.67
Scene 3	16.67	16.67	16.67	16.67	16.67	16.67	16.67	16.67	16.67	16.67
Scene 4	50.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
Scene 5	28.57	28.57	28.57	28.57	28.57	28.57	28.57	28.57	28.57	28.57
Scene 6	16.67	16.67	16.67	16.67	16.67	16.67	16.67	16.67	16.67	16.67
Scene 7	50.00	50.00	50.00	50.00	50.00	50.00	50.00	50.00	50.00	50.00
Scene 8	16.67	16.67	16.67	16.67	16.67	20.00	16.67	16.67	16.67	16.67
Scene 9	25.00	25.00	25.00	25.00	25.00	25.00	25.00	25.00	25.00	25.00
Scene 10	25.00	16.67	16.67	16.67	16.67	16.67	16.67	16.67	16.67	16.67

Image model of Scene 5 as compared image

	0°	10°	20°	30°	40°	50°	60°	70°	80°	90°
Scene 1	42.86	42.86	42.86	42.86	42.86	42.86	50.00	42.86	42.86	42.86
Scene 2	16.67	16.67	16.67	16.67	16.67	16.67	16.67	16.67	16.67	16.67
Scene 3	33.33	33.33	33.33	33.33	33.33	33.33	33.33	33.33	33.33	33.33
Scene 4	33.33	33.33	33.33	33.33	33.33	33.33	33.33	33.33	33.33	33.33
Scene 5	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
Scene 6	33.33	33.33	33.33	33.33	33.33	33.33	33.33	33.33	33.33	33.33
Scene 7	33.33	33.33	33.33	33.33	33.33	33.33	33.33	33.33	33.33	33.33
Scene 8	16.67	16.67	16.67	16.67	16.67	20.00	16.67	16.67	16.67	16.67
Scene 9	25.00	25.00	25.00	25.00	25.00	25.00	25.00	25.00	25.00	25.00
Scene 10	25.00	33.33	33.33	33.33	33.33	33.33	33.33	33.33	33.33	33.33

Image model of Scene 6 as compared image

	0°	10°	20°	30°	40°	50°	60°	70°	80°	90°
Scene 1	28.57	28.57	28.57	28.57	28.57	28.57	33.33	28.57	28.57	28.57
Scene 2	16.67	16.67	16.67	16.67	16.67	16.67	16.67	16.67	16.67	16.67
Scene 3	16.67	16.67	16.67	16.67	16.67	16.67	16.67	16.67	16.67	16.67
Scene 4	16.67	16.67	16.67	16.67	16.67	16.67	16.67	16.67	16.67	16.67
Scene 5	28.57	28.57	28.57	28.57	28.57	28.57	28.57	28.57	28.57	28.57
Scene 6	66.67	66.67	66.67	66.67	66.67	66.67	66.67	66.67	66.67	66.67
Scene 7	16.67	16.67	16.67	16.67	16.67	16.67	16.67	16.67	16.67	16.67
Scene 8	16.67	16.67	16.67	16.67	16.67	20.00	16.67	16.67	16.67	16.67
Scene 9	25.00	25.00	25.00	25.00	25.00	25.00	25.00	25.00	25.00	25.00
Scene 10	25.00	16.67	16.67	16.67	16.67	16.67	16.67	16.67	16.67	16.67

Image model of Scene 7 as compared image

	0°	10°	20°	30°	40°	50°	60°	70°	80°	90°
Scene 1	28.57	28.57	28.57	28.57	28.57	28.57	33.33	28.57	28.57	28.57
Scene 2	16.67	16.67	16.67	16.67	16.67	16.67	16.67	16.67	16.67	16.67
Scene 3	16.67	16.67	16.67	16.67	16.67	16.67	16.67	16.67	16.67	16.67
Scene 4	50.00	50.00	50.00	50.00	50.00	50.00	50.00	50.00	50.00	50.00
Scene 5	28.57	28.57	28.57	28.57	28.57	28.57	28.57	28.57	28.57	28.57
Scene 6	16.67	16.67	16.67	16.67	16.67	16.67	16.67	16.67	16.67	16.67
Scene 7	100.00	100.00	66.67	66.67	100.00	100.00	100.00	100.00	100.00	100.00
Scene 8	16.67	16.67	16.67	16.67	16.67	20.00	16.67	16.67	16.67	16.67
Scene 9	25.00	25.00	25.00	25.00	25.00	25.00	25.00	25.00	25.00	25.00
Scene 10	25.00	16.67	16.67	16.67	16.67	16.67	16.67	16.67	16.67	16.67

Image model of Scene 8 as compared image

	0°	10°	20°	30°	40°	50°	60°	70°	80°	90°
Scene 1	42.86	42.86	42.86	42.86	42.86	42.86	33.33	42.86	42.86	42.86
Scene 2	50.00	50.00	50.00	50.00	50.00	50.00	50.00	50.00	50.00	50.00
Scene 3	33.33	33.33	33.33	33.33	33.33	33.33	33.33	33.33	33.33	33.33
Scene 4	33.33	16.67	16.67	16.67	16.67	16.67	16.67	16.67	16.67	16.67
Scene 5	14.29	14.29	14.29	14.29	14.29	14.29	14.29	14.29	14.29	14.29
Scene 6	16.67	16.67	16.67	16.67	16.67	16.67	16.67	16.67	16.67	16.67
Scene 7	16.67	16.67	16.67	16.67	16.67	16.67	16.67	16.67	16.67	16.67
Scene 8	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
Scene 9	25.00	25.00	25.00	25.00	25.00	25.00	25.00	25.00	25.00	25.00
Scene 10	50.00	33.33	33.33	33.33	33.33	33.33	33.33	33.33	33.33	33.33

Image model of Scene 9 as compared image

	0°	10°	20°	30°	40°	50°	60°	70°	80°	90°
Scene 1	14.29	14.29	14.29	14.29	14.29	14.29	16.67	14.29	14.29	14.29
Scene 2	16.67	16.67	16.67	16.67	16.67	16.67	16.67	16.67	16.67	16.67
Scene 3	16.67	16.67	16.67	16.67	16.67	16.67	16.67	16.67	16.67	16.67
Scene 4	16.67	16.67	16.67	16.67	16.67	16.67	16.67	16.67	16.67	16.67
Scene 5	14.29	14.29	14.29	14.29	14.29	14.29	14.29	14.29	14.29	14.29
Scene 6	16.67	16.67	16.67	16.67	16.67	16.67	16.67	16.67	16.67	16.67
Scene 7	16.67	16.67	16.67	16.67	16.67	16.67	16.67	16.67	16.67	16.67
Scene 8	16.67	16.67	16.67	16.67	16.67	20.00	16.67	16.67	16.67	16.67
Scene 9	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
Scene 10	25.00	16.67	16.67	16.67	16.67	16.67	16.67	16.67	16.67	16.67

Image model of Scene 10 as compared image

	0°	10°	20°	30°	40°	50°	60°	70°	80°	90°
Scene 1	28.57	28.57	28.57	28.57	28.57	28.57	33.33	28.57	28.57	28.57
Scene 2	33.33	33.33	33.33	33.33	33.33	33.33	33.33	33.33	33.33	33.33
Scene 3	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
Scene 4	33.33	16.67	16.67	16.67	16.67	16.67	16.67	16.67	16.67	16.67
Scene 5	28.57	28.57	28.57	28.57	28.57	28.57	28.57	28.57	28.57	28.57
Scene 6	16.67	16.67	16.67	16.67	16.67	16.67	16.67	16.67	16.67	16.67
Scene 7	16.67	16.67	16.67	16.67	16.67	16.67	16.67	16.67	16.67	16.67
Scene 8	33.33	33.33	33.33	33.33	33.33	20.00	33.33	33.33	33.33	33.33
Scene 9	25.00	25.00	25.00	25.00	25.00	25.00	25.00	25.00	25.00	25.00
Scene 10	50.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00

3.) The results of similarity ratio between scene representative image at 0° and whole dataset using proposed algorithm

Image model of Scene 1 as compared image

	0°	10°	20°	30°	40°	50°	60°	70°	80°	90°
Scene 1	100.00	100.00	100.00	100.00	100.00	100.00	50.00	100.00	100.00	100.00
Scene 2	58.33	58.33	58.33	58.33	58.33	58.33	58.33	58.33	58.33	58.33
Scene 3	54.55	54.55	54.55	54.55	54.55	54.55	54.55	54.55	54.55	54.55
Scene 4	30.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Scene 5	61.54	61.54	61.54	61.54	61.54	61.54	61.54	61.54	61.54	61.54
Scene 6	20.00	20.00	20.00	20.00	20.00	20.00	20.00	20.00	20.00	20.00
Scene 7	18.18	18.18	18.18	18.18	18.18	18.18	18.18	18.18	18.18	18.18
Scene 8	41.67	41.67	41.67	41.67	41.67	33.33	41.67	41.67	41.67	41.67
Scene 9	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Scene 10	42.86	54.55	54.55	54.55	54.55	54.55	54.55	54.55	54.55	54.55

Image model of Scene 2 as compared image

	0°	10°	20°	30°	40°	50°	60°	70°	80°	90°
Scene 1	50.00	50.00	50.00	50.00	50.00	50.00	16.67	50.00	50.00	50.00
Scene 2	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
Scene 3	54.55	54.55	54.55	54.55	54.55	54.55	54.55	54.55	54.55	54.55
Scene 4	30.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Scene 5	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Scene 6	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Scene 7	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Scene 8	41.67	41.67	41.67	41.67	41.67	33.33	41.67	41.67	41.67	41.67
Scene 9	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Scene 10	42.86	54.55	54.55	54.55	54.55	54.55	54.55	54.55	54.55	54.55

Image model of Scene 3 as compared image

	0°	10°	20°	30°	40°	50°	60°	70°	80°	90°
Scene 1	42.86	42.86	42.86	42.86	42.86	42.86	33.33	42.86	42.86	42.86
Scene 2	50.00	50.00	50.00	50.00	50.00	50.00	50.00	50.00	50.00	50.00
Scene 3	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
Scene 4	30.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Scene 5	23.08	23.08	23.08	23.08	23.08	23.08	23.08	23.08	23.08	23.08
Scene 6	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Scene 7	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Scene 8	25.00	25.00	25.00	25.00	25.00	0.00	25.00	25.00	25.00	25.00
Scene 9	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Scene 10	42.86	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00

Image model of Scene 4 as compared image

	0°	10°	20°	30°	40°	50°	60°	70°	80°	90°
Scene 1	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Scene 2	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Scene 3	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Scene 4	80.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
Scene 5	23.08	23.08	23.08	23.08	23.08	23.08	23.08	23.08	23.08	23.08
Scene 6	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Scene 7	27.27	27.27	27.27	27.27	27.27	27.27	27.27	27.27	27.27	27.27
Scene 8	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Scene 9	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Scene 10	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

Image model of Scene 5 as compared image

	0°	10°	20°	30°	40°	50°	60°	70°	80°	90°
Scene 1	57.14	57.14	57.14	57.14	57.14	57.14	41.67	57.14	57.14	57.14
Scene 2	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Scene 3	27.27	27.27	27.27	27.27	27.27	27.27	27.27	27.27	27.27	27.27
Scene 4	30.00	27.27	27.27	27.27	27.27	27.27	27.27	27.27	27.27	27.27
Scene 5	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
Scene 6	40.00	40.00	40.00	40.00	40.00	40.00	40.00	40.00	40.00	40.00
Scene 7	36.36	36.36	36.36	36.36	36.36	36.36	36.36	36.36	36.36	36.36
Scene 8	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Scene 9	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Scene 10	0.00	27.27	27.27	27.27	27.27	27.27	27.27	27.27	27.27	27.27

Image model of Scene 6 as compared image

	0°	10°	20°	30°	40°	50°	60°	70°	80°	90°
Scene 1	14.29	14.29	14.29	14.29	14.29	14.29	25.00	14.29	14.29	14.29
Scene 2	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Scene 3	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Scene 4	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Scene 5	30.77	30.77	30.77	30.77	30.77	30.77	30.77	30.77	30.77	30.77
Scene 6	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
Scene 7	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Scene 8	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Scene 9	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Scene 10	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

Image model of Scene 7 as compared image

	0°	10°	20°	30°	40°	50°	60°	70°	80°	90°
Scene 1	21.43	21.43	21.43	21.43	21.43	21.43	25.00	21.43	21.43	21.43
Scene 2	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Scene 3	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Scene 4	40.00	27.27	27.27	27.27	27.27	27.27	27.27	27.27	27.27	27.27
Scene 5	38.46	38.46	38.46	38.46	38.46	38.46	38.46	38.46	38.46	38.46
Scene 6	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Scene 7	100.00	100.00	90.91	90.91	100.00	100.00	100.00	100.00	100.00	100.00
Scene 8	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Scene 9	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Scene 10	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

Image model of Scene 8 as compared image

	0°	10°	20°	30°	40°	50°	60°	70°	80°	90°
Scene 1	35.71	35.71	35.71	35.71	35.71	35.71	16.67	35.71	35.71	35.71
Scene 2	41.67	41.67	41.67	41.67	41.67	41.67	41.67	41.67	41.67	41.67
Scene 3	27.27	27.27	27.27	27.27	27.27	27.27	27.27	27.27	27.27	27.27
Scene 4	30.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Scene 5	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Scene 6	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Scene 7	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Scene 8	100.00	100.00	100.00	100.00	100.00	88.89	100.00	100.00	100.00	100.00
Scene 9	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Scene 10	42.86	27.27	27.27	27.27	27.27	27.27	27.27	27.27	27.27	27.27

Image model of Scene 9 as compared image

	0°	10°	20°	30°	40°	50°	60°	70°	80°	90°
Scene 1	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Scene 2	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Scene 3	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Scene 4	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Scene 5	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Scene 6	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Scene 7	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Scene 8	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Scene 9	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
Scene 10	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

Image model of Scene 10 as compared image

	0°	10°	20°	30°	40°	50°	60°	70°	80°	90°
Scene 1	42.86	42.86	42.86	42.86	42.86	42.86	33.33	42.86	42.86	42.86
Scene 2	50.00	50.00	50.00	50.00	50.00	50.00	50.00	50.00	50.00	50.00
Scene 3	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
Scene 4	30.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Scene 5	23.08	23.08	23.08	23.08	23.08	23.08	23.08	23.08	23.08	23.08
Scene 6	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Scene 7	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Scene 8	25.00	25.00	25.00	25.00	25.00	0.00	25.00	25.00	25.00	25.00
Scene 9	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Scene 10	42.86	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00

Time consuming result for single plane dataset

1.) The results of time consuming between scene representative image and the other image in the whole dataset using A* search algorithm in milliseconds (ms).

Image model of Scene 1 as compared image

	0°	10°	20°	30°	40°	50°	60°	70°	80°	90°
Scene 1	17.43	4.33	2.26	2.24	2.25	2.29	2.04	2.42	2.24	2.09
Scene 2	61.43	43.80	42.75	43.36	43.71	43.14	44.11	43.53	41.76	44.71
Scene 3	45.50	40.18	40.60	38.21	45.15	38.81	38.36	39.41	44.84	38.80
Scene 4	39.60	38.41	40.45	38.38	37.93	36.60	36.56	36.58	38.70	37.51
Scene 5	311.91	305.91	301.47	287.37	288.21	295.91	294.42	291.23	300.87	299.00
Scene 6	38.40	38.70	39.18	39.22	38.54	39.13	37.58	38.72	36.40	37.78
Scene 7	38.43	38.93	38.27	39.29	37.15	47.80	39.28	38.57	38.97	39.02
Scene 8	44.62	42.00	40.13	38.83	41.77	8.23	45.21	43.68	41.43	39.26
Scene 9	2.68	2.24	2.40	2.20	2.11	2.17	2.13	2.12	2.14	2.06
Scene 10	3.11	40.40	38.85	39.53	39.34	40.42	39.78	38.31	39.24	40.38

Image model of Scene 2 as compared image

	0°	10°	20°	30°	40°	50°	60°	70°	80°	90°
Scene 1	120.67	119.65	120.51	120.02	120.13	120.12	15.36	120.67	120.26	120.43
Scene 2	1.37	1.32	1.03	1.02	1.00	0.91	0.81	0.77	0.78	0.80
Scene 3	16.18	16.30	16.08	16.16	17.28	16.14	15.82	15.77	15.98	16.18
Scene 4	15.98	15.61	15.35	15.67	15.32	15.56	15.56	15.59	15.39	15.34
Scene 5	104.69	104.23	104.30	103.89	103.99	104.45	105.19	104.53	104.65	104.17
Scene 6	13.46	13.71	13.64	13.84	13.89	13.96	13.55	13.58	14.43	13.57
Scene 7	13.80	14.08	14.47	14.30	14.47	14.37	14.10	14.44	13.93	14.05
Scene 8	17.15	16.92	17.05	16.88	17.06	3.06	16.84	16.91	16.72	16.66
Scene 9	0.91	0.90	0.88	0.89	0.87	0.87	0.87	0.86	0.86	0.86
Scene 10	1.02	15.84	16.38	15.80	15.74	15.96	15.80	15.84	16.03	16.09

Image model of Scene 3 as compared image

	0°	10°	20°	30°	40°	50°	60°	70°	80°	90°
Scene 1	108.67	108.09	107.87	107.93	108.64	107.88	17.36	107.65	108.37	107.48
Scene 2	15.90	16.05	15.65	15.80	15.73	15.87	15.81	15.89	16.11	15.78
Scene 3	3.19	2.26	2.16	2.11	2.16	2.11	2.08	2.11	2.06	2.02
Scene 4	15.40	14.75	14.87	15.07	14.95	14.93	14.79	14.85	14.90	14.94
Scene 5	105.07	104.69	104.89	104.22	105.50	106.24	104.98	104.39	105.75	104.83
Scene 6	14.73	14.40	14.32	14.45	14.35	14.69	14.48	14.53	14.52	14.37
Scene 7	13.41	13.30	13.29	13.69	13.27	13.83	13.34	13.26	13.33	13.29
Scene 8	14.96	15.14	15.30	14.97	15.07	2.75	15.08	15.04	15.41	15.08
Scene 9	0.86	0.86	0.86	0.86	0.87	0.88	0.89	0.87	0.86	0.86
Scene 10	1.06	2.03	2.03	2.02	1.99	2.01	1.97	1.95	1.95	2.05

Image model of Scene 4 as compared image

	0°	10°	20°	30°	40°	50°	60°	70°	80°	90°
Scene 1	105.03	103.71	104.19	104.54	104.97	105.11	15.07	104.88	103.86	103.69
Scene 2	15.08	14.88	15.02	15.29	14.88	15.38	15.00	15.34	14.97	15.01
Scene 3	14.50	15.03	14.70	15.07	14.61	15.13	14.95	14.54	15.09	14.57
Scene 4	19.90	2.19	2.10	2.06	2.06	2.10	2.05	2.08	2.06	2.08
Scene 5	107.51	108.10	107.62	107.69	108.10	107.14	107.55	106.99	108.11	106.75
Scene 6	13.62	13.55	13.80	14.13	14.13	13.93	13.87	13.98	13.73	13.85
Scene 7	15.86	16.09	15.70	15.65	15.81	15.67	15.82	15.68	15.83	15.54
Scene 8	14.84	15.48	15.03	14.91	14.89	2.94	14.89	15.13	15.20	15.13
Scene 9	0.88	0.89	0.87	0.87	0.87	0.87	0.87	0.90	0.88	0.89
Scene 10	1.01	14.87	15.01	14.97	14.74	14.83	14.74	14.70	14.86	14.67

Image model of Scene 5 as compared image

	0°	10°	20°	30°	40°	50°	60°	70°	80°	90°
Scene 1	286.15	288.35	285.14	296.91	295.45	293.69	44.72	284.74	286.76	290.70
Scene 2	38.22	38.00	43.02	40.07	38.05	36.19	36.99	36.45	38.05	37.04
Scene 3	38.09	37.65	49.48	37.71	37.40	38.04	38.03	37.33	37.92	38.19
Scene 4	37.96	36.63	41.61	37.45	37.79	38.42	41.55	37.48	37.42	37.61
Scene 5	4.89	4.99	5.30	5.00	4.99	5.09	5.13	5.06	5.08	5.05
Scene 6	38.18	38.82	37.59	38.85	38.79	38.54	38.02	38.19	38.11	37.55
Scene 7	37.63	38.60	39.12	38.75	40.77	43.88	38.34	38.55	39.07	37.76
Scene 8	36.63	39.38	37.13	37.27	38.61	6.91	38.28	37.31	37.49	36.40
Scene 9	1.73	1.78	1.81	1.74	1.76	1.74	2.08	1.95	1.77	1.81
Scene 10	2.88	36.86	37.85	40.65	36.61	36.80	44.20	37.88	37.46	37.31

Image model of Scene 6 as compared image

	0°	10°	20°	30°	40°	50°	60°	70°	80°	90°
Scene 1	106.04	104.64	105.48	106.61	104.49	105.12	16.44	104.59	105.17	104.16
Scene 2	13.56	13.64	13.63	13.49	13.55	13.55	13.66	13.51	13.54	13.58
Scene 3	14.60	14.59	14.58	14.82	15.11	14.80	14.52	15.03	14.95	14.60
Scene 4	13.67	13.96	13.89	13.74	14.07	13.70	14.15	14.12	13.73	14.06
Scene 5	106.71	107.70	108.18	106.84	106.71	108.19	106.61	107.43	107.49	108.14
Scene 6	21.22	21.22	21.45	21.21	21.24	21.33	21.42	21.64	21.59	21.33
Scene 7	14.50	14.32	14.34	14.39	14.37	14.35	14.32	14.35	14.80	14.35
Scene 8	14.58	14.60	14.49	14.49	14.79	2.67	14.56	14.54	14.50	14.80
Scene 9	0.77	0.78	0.78	0.78	0.78	0.80	0.79	0.78	0.79	0.79
Scene 10	0.97	14.90	14.93	14.59	14.54	14.56	14.56	14.69	14.78	14.53

Image model of Scene 7 as compared image

	0°	10°	20°	30°	40°	50°	60°	70°	80°	90°
Scene 1	104.99	103.75	105.21	103.75	104.39	104.20	15.15	105.37	104.92	105.90
Scene 2	14.26	14.42	14.59	13.85	13.61	13.53	13.65	13.55	13.53	13.61
Scene 3	13.70	13.65	13.40	13.36	13.26	13.31	13.31	13.38	13.57	13.40
Scene 4	15.90	15.61	15.72	15.71	15.80	15.68	15.97	15.67	15.58	15.72
Scene 5	108.93	107.26	107.95	107.64	108.66	107.94	107.74	107.25	107.41	107.98
Scene 6	14.61	14.59	14.82	15.08	14.94	14.33	14.32	14.33	14.43	14.33
Scene 7	0.89	0.85	21.24	21.30	0.85	0.84	0.82	0.82	0.82	0.82
Scene 8	14.79	14.73	14.66	14.70	14.68	2.97	14.66	14.68	14.79	15.07
Scene 9	0.79	0.80	0.98	0.87	0.86	0.81	0.79	0.78	0.80	0.79
Scene 10	0.90	13.33	13.37	13.41	13.30	13.23	13.20	13.28	13.35	13.25

Image model of Scene 8 as compared image

	0°	10°	20°	30°	40°	50°	60°	70°	80°	90°
Scene 1	111.54	111.21	111.09	110.96	110.68	111.93	16.33	111.83	113.78	111.37
Scene 2	16.62	16.75	16.60	16.56	16.59	16.67	16.71	16.67	16.75	16.60
Scene 3	14.90	15.02	15.04	15.06	14.89	14.93	14.96	15.12	15.12	15.01
Scene 4	15.52	14.81	14.96	14.80	14.91	14.85	14.89	14.82	14.85	15.20
Scene 5	103.83	102.65	103.76	101.30	101.59	104.02	103.13	104.69	104.16	103.58
Scene 6	14.76	14.31	14.21	14.75	14.28	14.30	14.34	14.40	14.23	14.43
Scene 7	14.56	14.59	15.02	14.98	14.73	14.68	14.60	14.69	14.59	14.62
Scene 8	3.00	1.73	1.68	1.64	1.64	0.86	1.68	1.64	1.66	1.61
Scene 9	0.83	0.86	0.84	0.84	0.85	0.87	0.86	0.85	0.84	0.85
Scene 10	0.97	14.90	15.02	14.97	15.13	14.94	15.04	14.93	14.93	14.92

Image model of Scene 9 as compared image

	0°	10°	20°	30°	40°	50°	60°	70°	80°	90°
Scene 1	47.70	47.64	47.83	47.69	48.07	47.67	7.29	47.92	47.67	47.69
Scene 2	7.31	7.23	7.24	7.33	7.23	7.27	7.25	7.24	7.29	7.26
Scene 3	7.30	7.24	7.30	7.24	7.24	7.23	7.23	7.22	7.24	7.23
Scene 4	7.21	7.26	7.21	7.28	7.27	7.27	7.21	7.22	7.24	7.20
Scene 5	47.75	47.70	47.87	48.26	48.28	48.20	48.24	48.18	47.95	47.70
Scene 6	7.22	7.22	7.29	7.22	7.20	7.21	7.23	7.20	7.25	7.23
Scene 7	7.25	7.23	7.23	7.22	7.22	8.14	7.21	7.32	7.23	7.24
Scene 8	7.75	7.74	7.77	7.76	7.79	1.68	7.72	7.65	7.78	7.69
Scene 9	1.18	0.66	0.61	0.58	0.56	0.56	0.56	0.54	0.54	0.53
Scene 10	0.64	7.19	7.26	7.20	7.20	7.20	7.25	7.18	7.23	7.18

Image model of Scene 10 as compared image

	0°	10°	20°	30°	40°	50°	60°	70°	80°	90°
Scene 1	107.64	108.26	108.12	108.66	108.69	107.88	16.88	108.82	108.51	108.63
Scene 2	15.93	16.07	15.88	16.06	15.79	15.81	15.96	15.95	16.01	16.04
Scene 3	1.84	1.84	1.82	1.88	1.89	1.82	1.83	1.83	1.82	1.82
Scene 4	15.50	14.79	14.78	14.83	14.89	15.04	15.00	14.87	14.68	14.74
Scene 5	103.75	104.29	104.20	104.48	105.09	103.49	104.44	104.75	104.06	103.63
Scene 6	14.28	14.86	14.85	14.41	14.59	14.41	14.55	14.50	14.43	14.67
Scene 7	13.24	13.39	13.38	13.66	13.47	13.26	13.16	13.64	13.19	13.22
Scene 8	15.05	14.88	14.87	15.00	14.92	2.65	14.87	14.89	14.95	14.89
Scene 9	0.77	0.78	0.79	0.78	0.78	0.78	0.78	0.77	0.78	0.78
Scene 10	0.99	1.89	1.87	1.83	1.87	1.82	1.82	2.49	1.96	1.86

2.) The results of time consuming between scene representative image and the other image in the whole dataset using Decision tree Largest Common Sub-graph with tree model in milliseconds (ms).

Image model of Scene 1 as compared image

	0°	10°	20°	30°	40°	50°	60°	70°	80°	90°
Scene 1	29.60	7.21	7.37	6.03	17.21	3.39	80.22	3.84	2.38	2.22
Scene 2	16.62	12.45	11.84	11.70	12.92	12.16	11.86	11.36	11.31	11.11
Scene 3	10.63	9.74	9.45	9.62	9.93	9.57	9.61	9.50	9.58	9.42
Scene 4	9.31	8.87	9.23	9.01	8.95	8.89	8.99	8.97	8.79	9.01
Scene 5	338.25	344.18	344.68	340.94	350.08	339.97	339.33	341.48	342.21	339.03
Scene 6	8.78	8.72	8.57	8.67	8.69	8.81	8.58	8.64	8.81	8.64
Scene 7	8.74	8.75	9.04	8.82	8.95	8.79	8.72	8.82	8.71	8.83
Scene 8	9.57	10.48	9.84	9.89	10.03	2.78	10.15	9.97	10.19	9.83
Scene 9	11.04	2.08	0.92	0.98	0.94	0.84	0.90	0.82	0.88	0.80
Scene 10	4.24	9.64	9.57	9.25	9.85	9.51	9.33	9.34	9.26	9.57

Image model of Scene 2 as compared image

	0°	10°	20°	30°	40°	50°	60°	70°	80°	90°
Scene 1	11.05	6.57	6.41	6.38	6.86	6.49	1.49	6.25	6.35	6.27
Scene 2	1.13	1.09	0.96	0.91	0.88	0.89	0.80	0.80	0.81	0.80
Scene 3	1.56	1.59	1.53	1.54	1.50	1.48	1.48	1.48	1.48	1.50
Scene 4	1.49	1.34	1.38	1.30	1.30	1.29	1.31	1.29	1.33	1.40
Scene 5	4.53	4.55	4.53	4.52	4.47	4.45	4.55	4.49	4.49	4.49
Scene 6	0.69	0.64	0.67	0.64	0.63	0.64	0.63	0.63	0.64	0.63
Scene 7	0.64	0.67	0.63	0.63	0.64	0.73	0.66	0.67	0.65	0.64
Scene 8	1.64	1.73	1.86	1.86	1.82	0.88	1.86	1.76	1.71	1.73
Scene 9	0.87	0.53	0.49	0.46	0.43	0.43	0.43	0.40	0.40	0.40
Scene 10	2.34	1.87	1.49	1.57	1.49	1.47	1.47	1.47	1.47	1.47

Image model of Scene 3 as compared image

	0°	10°	20°	30°	40°	50°	60°	70°	80°	90°
Scene 1	4.99	4.99	4.96	4.91	4.94	4.95	4.94	4.95	4.96	4.89
Scene 2	1.52	1.48	1.47	1.49	1.47	1.48	1.47	1.53	1.45	1.47
Scene 3	1.69	1.70	1.68	1.62	1.61	1.65	1.60	1.88	1.59	1.58
Scene 4	1.40	1.28	1.45	1.45	1.37	1.42	1.31	1.28	1.27	1.27
Scene 5	4.53	4.57	4.50	4.50	4.55	4.51	4.51	4.50	4.55	4.51
Scene 6	1.20	1.37	1.24	1.21	1.22	1.21	1.19	1.19	1.19	1.19
Scene 7	0.66	0.69	0.67	0.65	0.63	0.63	0.64	0.63	0.62	0.63
Scene 8	1.34	1.35	1.35	1.33	1.35	0.61	1.32	1.32	1.34	1.33
Scene 9	0.57	0.43	0.40	0.43	0.41	0.41	0.46	0.41	0.41	0.40
Scene 10	0.95	1.60	1.60	1.55	1.56	1.63	1.71	1.65	1.66	1.63

Image model of Scene 4 as compared image

	0°	10°	20°	30°	40°	50°	60°	70°	80°	90°
Scene 1	4.66	4.69	4.61	4.62	4.63	4.73	1.33	4.62	4.67	4.72
Scene 2	1.33	1.32	1.32	1.34	1.33	1.33	1.41	1.41	1.44	1.42
Scene 3	1.36	1.36	1.35	1.36	1.41	1.37	1.30	1.36	1.33	1.33
Scene 4	27.23	2.91	2.80	2.87	2.90	2.84	2.83	2.81	2.88	2.88
Scene 5	4.96	4.97	5.07	4.86	4.75	5.39	4.87	4.78	4.81	4.82
Scene 6	0.64	0.82	0.69	0.64	0.63	0.66	0.63	0.62	0.63	0.63
Scene 7	0.51	0.55	0.49	0.50	0.49	0.49	0.49	0.58	0.52	0.55
Scene 8	1.37	1.37	1.34	1.36	1.46	0.72	1.34	1.34	1.33	1.28
Scene 9	0.53	0.43	0.42	0.42	0.40	0.40	0.40	0.39	0.40	0.39
Scene 10	0.90	1.26	1.35	1.27	1.25	1.25	1.24	1.25	1.25	1.24

Image model of Scene 5 as compared image

	0°	10°	20°	30°	40°	50°	60°	70°	80°	90°
Scene 1	343.35	346.11	345.73	346.10	343.17	342.70	51.11	344.42	344.38	349.88
Scene 2	8.97	9.09	8.75	8.85	8.78	8.94	8.61	8.94	8.66	8.81
Scene 3	8.86	8.71	8.77	8.57	8.63	9.06	9.12	8.63	9.13	8.50
Scene 4	8.86	9.09	9.03	9.23	9.12	9.11	9.47	9.44	9.04	9.04
Scene 5	6.05	5.95	5.93	6.26	6.03	5.92	6.09	6.69	6.40	6.12
Scene 6	8.98	9.05	8.87	9.58	8.77	8.78	8.82	8.96	8.91	8.98
Scene 7	9.00	9.20	9.21	9.18	8.97	8.94	9.15	8.97	9.15	9.02
Scene 8	8.70	8.68	8.44	8.56	8.48	2.33	8.48	8.31	8.83	8.42
Scene 9	0.78	0.62	0.59	0.55	0.54	0.54	0.53	0.55	0.58	0.54
Scene 10	2.42	8.76	8.91	8.84	8.73	8.62	8.62	8.57	8.55	8.75

Image model of Scene 6 as compared image

	0°	10°	20°	30°	40°	50°	60°	70°	80°	90°
Scene 1	4.62	4.60	4.57	4.60	4.54	4.73	4.62	4.51	4.51	4.61
Scene 2	0.62	0.61	0.59	0.60	0.60	0.59	0.59	0.60	0.61	0.60
Scene 3	1.20	1.21	1.18	1.19	1.19	1.19	1.18	1.19	1.20	1.19
Scene 4	0.64	0.66	0.65	0.63	0.61	0.60	0.61	0.61	0.61	0.61
Scene 5	4.80	4.90	4.79	4.82	4.89	4.83	4.83	4.85	4.83	4.80
Scene 6	29.60	29.43	29.58	29.51	29.37	29.35	29.52	29.41	29.59	29.48
Scene 7	1.24	1.19	1.18	1.17	1.16	1.16	1.19	1.26	1.27	1.25
Scene 8	1.26	1.23	1.26	1.25	1.26	0.62	1.23	1.25	1.23	1.24
Scene 9	0.46	0.38	0.37	0.37	0.36	0.37	0.37	0.37	0.37	0.37
Scene 10	1.35	1.26	1.27	1.26	1.29	1.26	1.23	1.23	1.21	1.23

Image model of Scene 7 as compared image

	0°	10°	20°	30°	40°	50°	60°	70°	80°	90°
Scene 1	4.73	4.75	4.62	4.61	4.71	4.74	1.32	4.64	4.69	4.66
Scene 2	0.65	0.63	0.62	0.62	0.62	0.68	0.63	0.62	0.62	0.63
Scene 3	0.60	0.61	0.60	0.60	0.60	0.60	0.60	0.59	0.60	0.59
Scene 4	0.51	0.51	0.48	0.49	0.47	0.47	0.48	0.47	0.49	0.48
Scene 5	4.97	4.98	4.97	4.94	5.05	4.97	4.92	5.16	4.98	4.93
Scene 6	1.16	1.20	1.18	1.16	1.15	1.16	1.18	1.17	1.15	1.14
Scene 7	1.63	1.66	30.04	30.24	1.63	1.72	1.61	1.59	1.58	1.56
Scene 8	1.24	1.24	1.23	1.29	1.24	1.24	1.32	1.23	1.23	1.22
Scene 9	0.42	0.40	0.38	0.37	0.37	0.39	0.38	0.38	0.37	0.37
Scene 10	0.67	0.62	0.61	0.61	0.63	0.60	0.61	0.60	0.60	0.60

Image model of Scene 8 as compared image

	0°	10°	20°	30°	40°	50°	60°	70°	80°	90°
Scene 1	5.35	5.33	5.25	5.29	5.33	5.32	1.36	5.29	5.22	5.27
Scene 2	1.61	1.60	1.59	1.60	1.59	1.62	1.61	1.60	1.62	1.64
Scene 3	1.31	1.27	1.28	1.34	1.29	1.34	1.29	1.28	1.26	1.27
Scene 4	1.39	1.26	1.27	1.75	1.46	1.33	1.32	1.39	1.46	1.41
Scene 5	4.97	4.60	4.56	4.66	4.43	4.40	4.64	4.45	4.41	4.43
Scene 6	1.29	1.18	1.16	1.19	1.15	1.14	1.14	1.14	1.17	1.17
Scene 7	1.25	1.21	1.20	1.22	1.21	1.21	1.21	1.22	1.21	1.25
Scene 8	0.69	0.69	0.65	0.65	0.66	0.65	0.66	0.66	0.65	0.65
Scene 9	0.53	0.50	0.47	0.51	0.45	0.44	0.44	0.45	0.44	0.45
Scene 10	0.68	1.30	1.28	1.28	1.27	1.27	1.37	1.31	1.29	1.27

Image model of Scene 9 as compared image

	0°	10°	20°	30°	40°	50°	60°	70°	80°	90°
Scene 1	4.21	0.62	0.43	0.41	0.40	0.42	0.37	0.37	0.36	0.37
Scene 2	0.35	0.37	0.35	0.35	0.35	0.35	0.35	0.36	0.35	0.37
Scene 3	0.35	0.35	0.35	0.35	0.35	0.34	0.34	0.34	0.34	0.34
Scene 4	0.35	0.34	0.36	0.37	0.36	0.36	0.41	0.36	0.36	0.58
Scene 5	0.35	0.35	0.34	0.34	0.34	0.34	0.34	0.34	0.34	0.34
Scene 6	0.33	0.34	0.34	0.34	0.35	0.34	0.34	0.34	0.36	0.34
Scene 7	0.33	0.34	0.35	0.34	0.35	0.35	0.35	0.35	0.36	0.46
Scene 8	1.99	0.54	0.48	0.47	0.50	2.35	0.47	0.44	0.44	0.43
Scene 9	0.55	0.53	0.50	0.49	0.48	0.48	0.49	0.47	0.47	0.47
Scene 10	1.75	0.38	0.35	0.35	0.35	0.35	0.35	0.35	0.35	0.35

Image model of Scene 10 as compared image

	0°	10°	20°	30°	40°	50°	60°	70°	80°	90°
Scene 1	5.05	5.02	5.00	4.91	4.87	4.96	4.91	4.85	4.91	4.89
Scene 2	1.43	1.43	1.43	1.48	1.43	1.42	1.49	1.43	1.42	1.43
Scene 3	1.47	1.51	1.46	1.47	1.46	1.46	1.46	1.45	1.46	1.46
Scene 4	1.39	1.23	1.31	1.22	1.22	1.22	1.21	1.22	1.21	1.22
Scene 5	4.48	4.48	4.44	4.46	4.49	4.43	4.45	4.46	4.45	4.44
Scene 6	1.22	1.16	1.15	1.23	1.16	1.16	1.14	1.15	1.15	1.17
Scene 7	0.62	0.59	0.59	0.59	0.60	0.59	0.60	0.59	0.59	0.59
Scene 8	1.29	1.31	1.31	1.34	1.30	0.60	1.31	1.30	1.29	1.29
Scene 9	0.39	0.39	0.37	0.36	0.37	0.37	0.36	0.36	0.37	0.37
Scene 10	0.79	1.52	1.50	1.47	1.46	1.46	1.51	1.51	1.46	1.47

3.) The results of time consuming between scene representative image and the other image in the whole dataset using proposed algorithm in milliseconds (ms).

Image model of Scene 1 as compared image

	0°	10°	20°	30°	40°	50°	60°	70°	80°	90°
Scene 1	27.18	4.02	3.06	3.04	3.42	3.32	22.02	4.64	3.13	2.79
Scene 2	4.99	3.58	3.11	2.79	2.99	2.98	2.78	2.69	2.54	2.48
Scene 3	10.37	2.86	2.53	2.50	2.67	2.35	2.33	2.24	2.18	2.11
Scene 4	5.49	1.76	2.77	2.31	2.21	2.03	1.64	1.53	1.51	1.26
Scene 5	6.03	3.04	2.79	2.68	2.57	2.08	1.98	2.07	2.34	1.92
Scene 6	7.39	4.31	2.74	2.86	3.17	2.85	2.75	2.57	2.61	2.47
Scene 7	3.27	2.49	3.80	3.34	2.91	2.38	2.09	1.97	1.92	1.82
Scene 8	4.84	3.75	2.71	2.47	2.17	3.19	2.25	2.09	2.12	1.96
Scene 9	1.85	1.96	1.11	1.27	1.04	1.00	0.92	0.97	1.02	0.86
Scene 10	5.04	2.91	2.00	1.81	1.63	1.55	1.38	1.26	1.39	1.20

Image model of Scene 2 as compared image

	0°	10°	20°	30°	40°	50°	60°	70°	80°	90°
Scene 1	1.62	1.38	1.09	1.05	1.00	1.10	2.21	2.91	1.00	0.86
Scene 2	1.51	2.30	1.70	1.47	1.18	1.08	1.05	1.06	1.09	1.03
Scene 3	2.16	0.74	0.71	0.71	0.71	0.71	0.71	0.71	0.70	0.73
Scene 4	2.29	1.44	0.87	0.63	0.62	0.60	0.61	1.17	1.03	0.84
Scene 5	1.11	1.95	1.12	1.01	1.08	1.22	0.99	1.03	0.58	0.47
Scene 6	1.22	0.78	0.70	0.68	0.60	0.74	0.64	0.43	0.42	0.51
Scene 7	0.72	0.45	0.40	0.42	0.39	0.42	0.58	0.44	0.38	0.38
Scene 8	0.74	1.62	1.28	1.19	0.87	1.38	0.78	0.71	0.67	0.67
Scene 9	0.71	0.56	0.48	0.51	0.45	0.40	0.40	0.41	0.39	0.35
Scene 10	1.59	1.78	0.76	0.71	0.71	0.71	0.73	0.72	0.72	0.71

Image model of Scene 3 as compared image

	0°	10°	20°	30°	40°	50°	60°	70°	80°	90°
Scene 1	8.37	1.40	1.27	1.32	1.30	1.44	3.23	2.12	1.36	1.33
Scene 2	1.73	1.12	1.20	1.10	1.03	0.98	0.97	0.93	0.84	0.81
Scene 3	5.71	1.68	1.62	1.63	1.49	1.51	1.49	1.51	1.54	1.69
Scene 4	6.02	5.24	2.68	1.50	1.51	1.28	1.32	1.35	1.28	1.13
Scene 5	6.15	1.24	1.13	1.18	1.07	1.04	1.06	1.17	1.20	1.19
Scene 6	3.50	0.94	0.89	0.90	0.84	0.73	0.72	0.67	0.69	0.72
Scene 7	1.84	0.68	1.30	0.91	0.58	0.56	0.54	0.53	0.50	0.47
Scene 8	3.23	1.36	1.18	1.04	0.87	1.27	0.87	0.74	0.73	0.69
Scene 9	0.40	0.39	0.35	0.37	0.34	0.34	0.34	0.34	0.34	0.34
Scene 10	3.45	4.75	1.61	1.49	1.29	1.29	1.21	1.05	1.06	1.02

Image model of Scene 4 as compared image

	0°	10°	20°	30°	40°	50°	60°	70°	80°	90°
Scene 1	3.10	3.48	2.06	1.86	1.46	1.40	2.35	1.26	1.03	0.89
Scene 2	1.09	2.34	1.10	1.01	0.80	0.71	0.67	0.68	0.65	0.57
Scene 3	2.72	3.63	1.40	1.39	1.03	1.19	1.07	1.00	0.85	0.95
Scene 4	8.40	2.53	2.29	1.36	1.28	1.21	1.24	1.18	1.17	1.23
Scene 5	6.19	3.41	1.92	1.80	1.77	1.53	1.38	1.35	1.35	1.16
Scene 6	3.53	0.90	0.81	0.67	0.60	0.55	0.44	0.44	0.44	0.41
Scene 7	5.83	3.37	1.64	1.48	1.47	1.37	1.32	1.29	1.20	1.28
Scene 8	1.54	2.01	0.67	0.60	0.60	1.29	0.59	0.55	0.54	0.57
Scene 9	0.67	0.55	0.58	0.54	0.45	0.50	0.43	0.43	0.41	0.40
Scene 10	0.82	0.94	0.83	0.70	0.64	0.59	0.57	0.57	0.62	0.57

Image model of Scene 5 as compared image

	0°	10°	20°	30°	40°	50°	60°	70°	80°	90°
Scene 1	7.16	2.72	1.69	1.64	1.54	1.54	5.77	4.04	1.45	1.46
Scene 2	0.96	0.96	0.90	0.83	0.74	0.71	0.63	0.61	0.64	0.62
Scene 3	2.73	1.40	1.46	1.34	1.25	1.03	0.95	0.99	1.09	1.01
Scene 4	5.87	1.71	1.57	1.34	1.34	1.37	1.37	1.35	1.21	1.11
Scene 5	5.20	1.57	1.54	1.48	1.46	1.46	1.47	1.42	1.41	1.41
Scene 6	6.41	2.75	1.97	1.98	1.97	1.76	1.68	1.66	1.61	1.52
Scene 7	6.65	2.83	3.53	2.55	2.17	1.80	1.66	1.59	1.64	1.43
Scene 8	2.55	2.89	1.89	1.85	1.41	3.62	1.45	1.08	0.90	0.85
Scene 9	2.59	2.19	1.18	1.14	0.89	0.95	0.94	0.99	0.81	0.78
Scene 10	5.19	4.47	1.13	1.00	0.82	0.80	0.86	0.81	0.81	0.81

Image model of Scene 6 as compared image

	0°	10°	20°	30°	40°	50°	60°	70°	80°	90°
Scene 1	3.46	4.59	2.25	1.97	1.77	1.74	2.86	3.44	1.38	1.27
Scene 2	1.96	2.76	1.21	1.14	1.06	1.01	0.89	0.63	0.55	0.60
Scene 3	2.69	0.87	0.71	0.75	0.57	0.52	0.45	0.45	0.47	0.43
Scene 4	3.45	1.11	0.92	0.86	0.83	0.79	0.72	0.67	0.65	0.61
Scene 5	5.07	2.42	2.11	2.17	1.97	1.39	1.42	1.32	1.18	1.09
Scene 6	1.39	4.25	1.78	1.61	1.63	1.77	1.49	1.23	1.14	1.12
Scene 7	2.86	2.96	1.87	1.88	0.85	1.08	0.85	0.76	0.64	0.58
Scene 8	1.69	0.57	0.58	0.52	0.45	0.41	0.40	0.41	0.36	0.33
Scene 9	2.49	0.87	0.66	0.55	0.42	0.34	0.36	0.37	0.31	0.31
Scene 10	1.18	1.89	1.01	0.93	0.70	0.63	0.59	0.43	0.47	0.44

Image model of Scene 7 as compared image

	0°	10°	20°	30°	40°	50°	60°	70°	80°	90°
Scene 1	2.28	1.37	0.88	0.93	1.02	0.71	2.18	0.60	0.56	0.55
Scene 2	0.32	0.43	0.36	0.34	0.32	0.32	0.32	0.32	0.31	0.32
Scene 3	0.69	0.69	0.49	0.48	0.39	0.38	0.36	0.36	0.33	0.36
Scene 4	7.33	2.42	1.41	1.27	1.12	1.00	0.97	0.87	0.82	0.86
Scene 5	2.94	2.01	1.55	1.30	1.12	1.07	0.99	0.83	0.80	0.76
Scene 6	0.80	2.20	1.15	1.02	0.96	0.70	0.60	0.74	0.70	0.72
Scene 7	3.31	2.06	3.41	1.34	1.00	0.99	0.96	0.93	0.97	1.05
Scene 8	0.73	0.52	0.36	0.32	0.32	0.32	0.32	0.32	0.34	0.32
Scene 9	1.79	0.36	0.29	0.29	0.29	0.29	0.30	0.29	0.29	0.28
Scene 10	1.37	1.31	0.36	0.32	0.35	0.32	0.32	0.32	0.32	0.32

Image model of Scene 8 as compared image

	0°	10°	20°	30°	40°	50°	60°	70°	80°	90°
Scene 1	1.16	0.81	0.65	0.63	0.62	0.75	2.38	1.02	0.65	0.64
Scene 2	1.76	0.82	0.69	0.67	0.66	0.65	0.66	0.65	0.65	0.62
Scene 3	2.34	0.92	0.63	0.72	0.65	0.58	0.58	0.62	0.53	0.50
Scene 4	5.20	0.90	0.61	0.50	0.46	0.41	0.37	0.36	0.33	0.33
Scene 5	0.35	0.35	0.35	0.34	0.34	0.34	0.34	0.34	0.38	0.34
Scene 6	0.32	1.16	0.40	0.37	0.43	0.38	0.36	0.33	0.33	0.33
Scene 7	0.33	0.33	0.33	0.35	0.35	0.34	0.34	0.33	0.33	0.33
Scene 8	0.95	0.90	0.88	0.90	0.89	3.66	0.90	0.94	0.88	0.87
Scene 9	1.07	0.48	0.45	0.33	0.30	0.31	0.30	0.30	0.30	0.30
Scene 10	3.38	4.02	0.55	0.51	0.50	0.49	0.52	0.53	0.51	0.50

Image model of Scene 9 as compared image

	0°	10°	20°	30°	40°	50°	60°	70°	80°	90°
Scene 1	1.58	0.45	0.47	0.36	0.37	0.41	0.34	0.34	0.34	0.32
Scene 2	0.33	0.33	0.33	0.31	0.31	0.32	0.31	0.28	0.28	0.28
Scene 3	0.27	0.28	0.28	0.28	0.28	0.28	0.28	0.28	0.28	0.28
Scene 4	0.29	0.28	0.29	0.29	0.29	0.28	0.28	0.28	0.28	0.28
Scene 5	0.48	0.86	0.50	0.42	0.35	0.37	0.33	0.36	0.35	0.31
Scene 6	0.30	0.31	0.31	0.31	0.38	0.31	0.31	0.31	0.32	0.30
Scene 7	0.35	0.30	0.29	0.29	0.29	0.28	0.28	0.28	0.28	0.28
Scene 8	0.30	1.42	0.50	0.44	0.35	1.98	0.36	0.35	0.35	0.35
Scene 9	2.65	0.85	0.81	0.78	0.74	0.72	0.82	0.76	0.73	0.72
Scene 10	0.51	0.30	0.29	0.29	0.28	0.28	0.28	0.29	0.29	0.29

Image model of Scene 10 as compared image

	0°	10°	20°	30°	40°	50°	60°	70°	80°	90°
Scene 1	1.17	1.14	0.92	0.88	0.74	0.75	2.17	1.27	0.67	0.71
Scene 2	0.79	0.65	0.68	0.66	0.65	0.65	0.64	0.65	0.64	0.64
Scene 3	0.91	0.95	0.92	0.89	0.87	0.87	0.90	0.87	0.87	0.86
Scene 4	4.41	1.35	0.98	0.49	0.41	0.41	0.40	0.40	0.37	0.37
Scene 5	1.60	1.05	0.80	0.69	0.62	0.53	0.53	0.55	0.54	0.52
Scene 6	0.35	0.43	0.33	0.32	0.32	0.32	0.32	0.32	0.32	0.32
Scene 7	0.63	0.57	0.50	0.43	0.36	0.34	0.33	0.33	0.32	0.32
Scene 8	0.69	0.62	0.52	0.51	0.50	0.93	0.52	0.51	0.53	0.50
Scene 9	0.28	0.34	0.29	0.28	0.28	0.28	0.29	0.28	0.28	0.28
Scene 10	2.71	2.61	0.91	0.86	0.86	0.87	0.86	0.88	0.87	0.88

REFERENCES

1. Brezeale, D. and D.J. Cook, *Automatic Video Classification: A Survey of the Literature.* Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on, 2008. **38**(3): p. 416-430.
2. Jones, S. and L. Shao, *Content-based retrieval of human actions from realistic video databases.* Information Sciences, 2013. **236**: p. 56-65.
3. Ding, G., J. Wang, and K. Qin, *A visual word weighting scheme based on emerging itemsets for video annotation.* Inf. Process. Lett., 2010. **110**(16): p. 692-696.
4. Tianzhu, Z., et al., *A Generic Framework for Video Annotation via Semi-Supervised Learning.* Multimedia, IEEE Transactions on, 2012. **14**(4): p. 1206-1219.
5. Conte, D., et al., *Thirty Years of graph Matching in Pattern Recognition.* International Journal of Pattern Recognition and Artificial Intelligence, 2004. **18**(03): p. 265-298.
6. Pawar, V.S. and M.A. Zaveri. *Graph based pattern matching*. in *Fuzzy Systems and Knowledge Discovery (FSKD), 2011 Eighth International Conference on*. 2011.
7. Blondel, V.D., et al., *A Measure of Similarity between Graph Vertices: Applications to Synonym Extraction and Web Searching.* SIAM Rev., 2004. **46**(4): p. 647-666.
8. Zager, L.A. and G.C. Verghese, *Graph similarity scoring and matching.* Applied Mathematics Letters, 2008. **21**(1): p. 86-94.
9. Chevalier, F., et al., *Retrieval of objects in video by similarity based on graph matching.* Pattern Recogn. Lett., 2007. **28**(8): p. 939-949.
10. Cao, Z. and M. Zhu, *An efficient video similarity search algorithm.* IEEE Transactions on Consumer Electronics, 2010. **56**(2): p. 751-755.
11. Gang, L. and W. Xiaochi, *Improved Bags-of-Words Algorithm for Scene Recognition.* Physics Procedia, 2012. **24, Part B**: p. 1255-1261.
12. Su, J.-H., et al., *Effective content-based video retrieval using pattern-indexing and matching techniques.* Expert Systems with Applications, 2010. **37**(7): p. 5068-5085.
13. Jin-fu, Y., et al. *Research on object recognition using bag of word model for mobile robot navigation*. in *Mechatronics and Automation (ICMA), 2011 International Conference on*. 2011.
14. Shearer, K., H. Bunke, and S. Venkatesh, *Video indexing and similarity retrieval by largest common subgraph detection using decision trees.* Pattern Recognition, 2001. **34**(5): p. 1075-1091.
15. Shearer, K., S. Venkatesh, and D. Kieronska, *Spatial Indexing for Video Databases.* Journal of Visual Communication and Image Representation, 1996. **7**(4): p. 325-335.
16. Arndt, T. and C. Shi-Kuo. *Image sequence compression by iconic indexing*. in *Visual Languages, 1989., IEEE Workshop on*. 1989.
17. Singh, R., J. Xu, and B. Berger, *Global alignment of multiple protein interaction networks with application to functional orthology detection.*

Proceedings of the National Academy of Sciences, 2008. **105**(35): p. 12763-12768.
18. Lowe, D.G. *Object recognition from local scale-invariant features*. in *Computer Vision, 1999. The Proceedings of the Seventh IEEE International Conference on*. 1999.
19. Lowe, D.G., *Distinctive Image Features from Scale-Invariant Keypoints.* Int. J. Comput. Vision, 2004. **60**(2): p. 91-110.
20. Mekhalfi, M.L., et al., *Toward an assisted indoor scene perception for blind people with image multilabeling strategies.* Expert Syst. Appl., 2015. **42**(6): p. 2907-2918.
21. Shih-Wei, S., et al. *Automatic annotation of Web videos*. in *Multimedia and Expo (ICME), 2011 IEEE International Conference on*. 2011.
22. Li-Qun, X. and L. Yongmin. *Video classification using spatial-temporal features and PCA*. in *Multimedia and Expo, 2003. ICME '03. Proceedings. 2003 International Conference on*. 2003.
23. Feng, Z. and F. De la Torre. *Factorized graph matching*. in *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*. 2012.
24. Lee, J., J. Oh, and S. Hwang, *Scenario based dynamic video abstractions using graph matching*, in *Proceedings of the 13th annual ACM international conference on Multimedia*. 2005, ACM: Hilton, Singapore. p. 810-819.
25. Dahm, N., et al., *Efficient subgraph matching using topological node feature constraints.* Pattern Recognition, 2015. **48**(2): p. 317-330.
26. Bao, S.Y., M. Sun, and S. Savarese, *Toward coherent object detection and scene layout understanding.* Image and Vision Computing, 2011. **29**(9): p. 569-579.
27. Ramalingam, S., et al. *Manhattan Junction Catalogue for Spatial Reasoning of Indoor Scenes*. in *2013 IEEE Conference on Computer Vision and Pattern Recognition*. 2013.
28. Dickinson, A.S.a.S., *Applications of Bipartite Matching to Problems in Object Recognition*, in *In Proceedings, ICCV Workshop on Graph Algorithms and Computer Vision*. 1999.
29. Serratosa, F., *Fast computation of Bipartite graph matching.* Pattern Recognition Letters, 2014. **45**: p. 244-250.
30. Su, H., et al. *Learning a dense multi-view representation for detection, viewpoint classification and synthesis of object categories*. in *2009 IEEE 12th International Conference on Computer Vision*. 2009.
31. Shoaib, M., et al., *Estimating layout of cluttered indoor scenes using trajectory-based priors.* Image and Vision Computing, 2014. **32**(11): p. 870-883.
32. Hödlmoser, M. and B. Micusik, *Surface Layout Estimation Using Multiple Segmentation Methods and 3D Reasoning.* 2013: p. 41-49.
33. Allen, J.F. and L.F. Allen, *Maintaining knowledge about temporal intervals.* Communication of ACM, 1983: p. 832-843.
34. Quattoni, A.a.T., A. *Recognizing indoor scenes*. in *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*. 2009.
35. contributors, W. *HSL and HSV --- Wikipedia{,} The Free Encyclopedia*. 2017; Available from: https://en.wikipedia.org/wiki/HSL_and_HSV.

VITA

Suprachaya Veeraprasit graduated Bachelor degree from Mahidol University in Information and Communication Technology Program, in 2008 academic year.

She graduated Master degree from Chulalongkorn University in Computer Science and Information Technology Program, in 2010 academic year.