

# Extended Spatio-temporal Relations between moving and Non-moving objects

Chang Choi<sup>1</sup>, Junho Choi<sup>1</sup>, Juhyun Shin<sup>1</sup>, Ilsun You<sup>2</sup> and Pankoo Kim<sup>3</sup>

<sup>1</sup> Dept. of Computer Engineering, Chosun University  
375 Seosuk-dong Donggu, Gwangju 501-759, South Korea  
enduranceaura@gmail.com, xdman@paran.com, jhshin@chosun.ac.kr

<sup>2</sup> School of Information Science, Korean Bible University  
16 Danghyun 2-gil, Nowon-gu, Seoul 139-791, South Korea  
isyou@bible.ac.kr

<sup>3</sup> Corresponding author, Dept. of CE, Chosun University, South Korea  
pkkim@chosun.ac.kr

**Abstract.** During the last decade, the emerging technology for video retrieval is mainly based on the content. However, semantic-based video retrieval has become more and more necessary for the humans especially the naive users who can only use the human language during retrieval. In this paper, we focus on semantic representation using topological and directional relations between non-moving and moving objects for security using CCTV(closed-circuit Television). In this paper, we propose new spatio-temporal relation to extend previous work using topological and directional relations and investigate spatio-temporal predicates which propose our models. In the experiment part, we compared retrieval results using TSR(Tangent Space Representation) with those using rules represented by the proposed model.

**Keywords:** Spatio-temporal Relation, Motion Recognition, Motion Predicates

## 1 Introduction

During the last decade, the emerging technology for video retrieval is mainly based on the content. However, semantic-based video retrieval has become more and more necessary for the humans especially the naive users who can only use the human language during retrieval. So, semantic-based video retrieval research has caused many researchers' attentions.

Since the most important semantic information for video is based on video motion research which is the significant factor for video event representation, there has been a significant amount of event understanding research in various application domains[12]. One major goal of this research is to accomplish the automatic extraction of feature semantics from a motion and to provide support for semantic-based motion indexing, retrieval and management. Most of the current approaches to activity recognition are composed of defining models for specific activity types that suit the goal in a particular domain and developing procedural recognized by

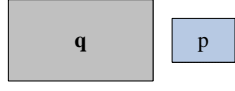
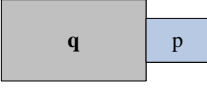

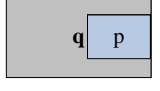
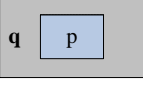

constructing the dynamic models of the periodic pattern of human movements and are highly dependent on the robustness of the tracking.

Spatio-temporal relations are the basis for many of the selections users perform when they formulate queries for the purpose of semantic-based motion retrieval. Although such query languages use natural-language-like terms, the formal definitions of these relations rarely reflect the language people would use when communicating with each other. To bridge the gap between the computational models used for spatio-temporal relations and people's use of motion verbs in their natural language, a model of these spatio-temporal relations was calibrated for motion verbs.

In the previous works, the retrieval using spatio-temporal relations is similar trajectory retrieval, it's only the content-based retrieval but not semantic-based. Therefore, we propose a novel approach for motion recognition from the aspect of semantic meaning in this paper. This issue can be addressed through a hierarchical model that explains how the human language interacts with motions and we evaluate our new approach using trajectory distance based on spatial relations to distinguish the conceptual similarity and get the satisfactory results. In the experiment and application part, we apply the proposed approach to semantic recognition of motions and trajectory retrieval. Extending our novel motion verbs model with more abundant motion verbs for gapping the chasm between high-level semantics and low-level video feature is our further consideration. Finally, we can apply a security using CCTV through this study

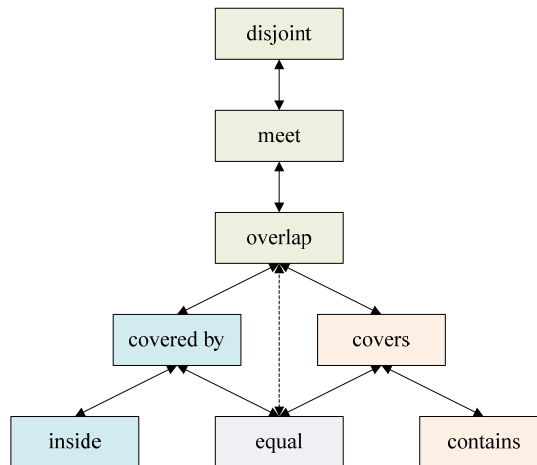
## 2 Related Works

The trajectory of a moving object is a presentation of spatio-temporal relationship between moving objects and an important element in video indexing for content-based retrieval. In Temporal relation, John Z. Li et al.[14] represented the trajectory of a moving object as eight directions – North(NT), Northwest(NW), Northeast(NE), West(WT), Southwest(SW), East(ET), Southeast(SE) and Southwest(SW). There are many researches on spatial relations of moving objects. Martin Erwig[2] is defined spatio-temporal predicates and Our previous work[3] was proposed Hierarchical relation modeling about moving object. Lee[4][13] have proposed relations such as 3D C-string for the various extensions of the original representations of 2D and W Ren, M Singh and S Singh[7] proposed to combine directional relations with topological relations by specifying six spatial relationships: 'Left', 'Right', 'Up', 'Down', 'Touch', and 'Front'. Pei-Yi Chen[8] measure velocity similarity by six possible velocity trends. The SMR scheme by Chang et al.[5][6] makes use of fifteen topological operators to describe the spatial relationships between moving object and non-moving object and also makes a new scheme to identify objects. Egnehofer and Franzosa[1] that proposed eight topological spatial relationships including 'equal', 'inside', 'cover', 'overlap', 'touch', 'disjoint', 'covered-by' and 'contains'. This model is the most famous study.

		
<b>disjoint (p , q)</b>	<b>meet (p , q)</b>	<b>overlap (p , q)</b>
		
<b>covered by (p , q) covers (q , p)</b>	<b>inside (p , q) contains (q , p)</b>	<b>equal (p , q)</b>

**Fig. 1.** Topological Relation of Egnehofer and Franzosa

For example, if two regions are disjoint and later they overlap, then there must have been a topological situation where both regions have met each other and they cannot move directly such as, from disjoint to equal in the fig. 2.



**Fig. 2.** Conceptual neighborhood graph

However, Most of study represents motion between 2 moving objects and some study is not sufficient to provide a complete representation of the semantic motion based on directional relations by ignoring topological relations and describing the concept or semantics motion. Therefore, we represent the semantics between moving object and non-moving object through topological relations that we make elements of motion. We also match motion verbs so that they don't have direction features for intrusion detection using CCTV[15][16].

### 3 Previous Work

In the previous work[3], we propose a novel approach for motion recognition from the aspect of semantic meaning. We define 5 basic elements to represent spatial relations between non-moving object and moving object, and then expand them through a hierarchical model. Therefore, it specifies how to give definitions to the motions of general moving objects by the spatio-temporal relation, and match between the definition of moving objects and motion verbs in the vocabulary.

The fig. 3 shows hierarchical semantic relations[11] for motion verbs, in other words, motion verbs and visual information map from low-level features to the semantic-level. Our model is divided into 3 levels. The level 1 describes basic elements of motion, level 2 is the extension of basic elements and level 3 shows combination of level 2. And, we represent semantic relation between motions and motion verbs for each element. For example, 'go into' and 'go out' are subclasses of motion word 'go through' which was set with 'Part\_of' relation and 'go\_to'('come up) can be inserted into 'go\_through', it also has a relation of antonym between 'go\_out'(depart) and 'enter'. Even if each motion verb is different in directional and spatial relations, the semantic relation is the same.

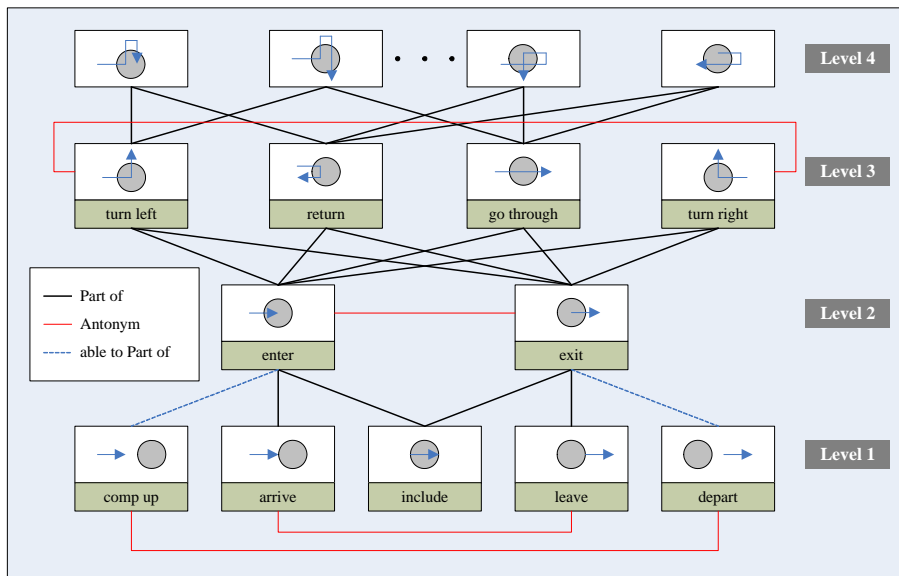


Fig. 3. Hierarchical relation modeling about moving object

However, it's difficult to distinguish semantics only using topological relations in level 3, which contains trajectories in the same topological relation. That is, we have to consider directional relations. We introduce proposed new spatio-temporal relation in next section and measure trajectory similarity using TSR which is represented moving distance and angle.

## 4 Building of Extended Spatio-temporal Relation

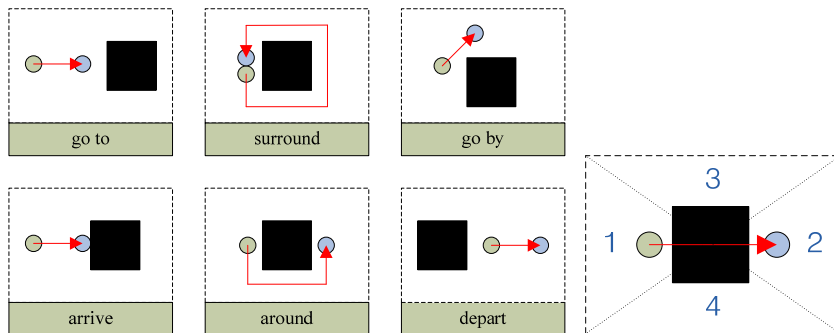
In this section we divided non-directional relation and directional relation.

### 4.1 Modeling of Non-Directional Relation Motion

In the fig. 3 is cannot representation of all of Motion so we extended our spatio-temporal relation using Egenhofer's model and our previous work. Firstly, we define motion using initial and final topological relation. For example, if initial topological relation is *Disjoint* and Final topological relation is *Disjoint* or *Meet* following fig. 2, table 1 is shown motion. For the motion, we considered the number of all cases.

**Table 1.** Representation of Topological Relation about *disjoint*

No.	Initial		Final		Motion
	Topology	position	Topology	position	
1	disjoint	1	disjoint	1	go to
2	disjoint	1	disjoint	1	surround
3	disjoint	1	disjoint	1	go by
4	disjoint	1	meet	1	arrive
5	disjoint	1	disjoint	2	around
6	disjoint	2	disjoint	2	depart
	disjoint	1	overlap	1	X
	disjoint	1	covered by	1	X
	disjoint	1	covers	1	X
	disjoint	1	inside	all	X
	disjoint	1	contains	1	X
	disjoint	1	equal	all	X



**Fig. 4.** Motion and Topological Position

Fig. 4 is shown motion and motion verbs also the position is shown in the table 1. Actually, *Disjoint* is can move the *disjoint* and *meet* but it cannot move the others such as *overlap*, *inside* and etc. from fig. 2. And *disjoint* can represent *go to*, *surround*,

*go by, arrive, around* and *depart* from initial and final position. Table 2 is the representation of topological relation about others.

**Table 2.** Representation of Topological Relation about others

No.	Initial		Final		Motion
	Topology	position	Topology	position	
7	meet	1	disjoint	1	leave
8	meet	1	meet	1	contact
9	meet	1	overlap	1	enter
10	overlap	1	meet	1	exit
11	overlap	1	overlap	1	contact
12	overlap	1	covered by	1	enter
13	overlap	1	covers	1	include
14	overlap	1	equal	1	same
15	covered by	1	overlap	1	exit
16	covered by	1	covered by	1	included by
17	covered by	1	inside	all	included by
18	covers	1	overlap	1	exit
19	covers	1	covers	1	include
20	covers	1	contains	1	include
21	inside	all	covered by	all	included by
22	inside	all	inside	all	included by
23	contains	all	covers	all	include
24	contains	all	contains	all	include
25	equal	-	overlap	all	exit
26	equal	-	equal	-	same

We get 14 motions from table 1 and 2. Some motion verb is including several motions. Even if initial position changed, it is not a problem because we consider all of one way direction. This means if initial position changed, final position also changed. Other case is combination of motion we can solve using combination and it made a new model such as fig. 5. For example DJ\_OL from table 3(For example, if initial topological relation is DJ and Final topological relation is also DJ, we'll represent DJ\_DJ) is can be representation using fig. 2 but we need this case so we combine DJ\_ME with ME\_OL. In this case 2DJ is a sufficient condition for DJ\_OL.

**Table 3.** Representation of Topological Relation

Representation	Topological Relation	Representation	Topological Relation
DJ	Disjoint	CD	Covered by
ME	Meet	CR	Covers
OL	Overlap	IN	Inside
EQ	Equal	CN	Contains

Finally, we define 51 motions it doesn't consider object size. It means that some motion represent covered by or covers between moving object size and non-moving object size so except this case we get 51 motions.

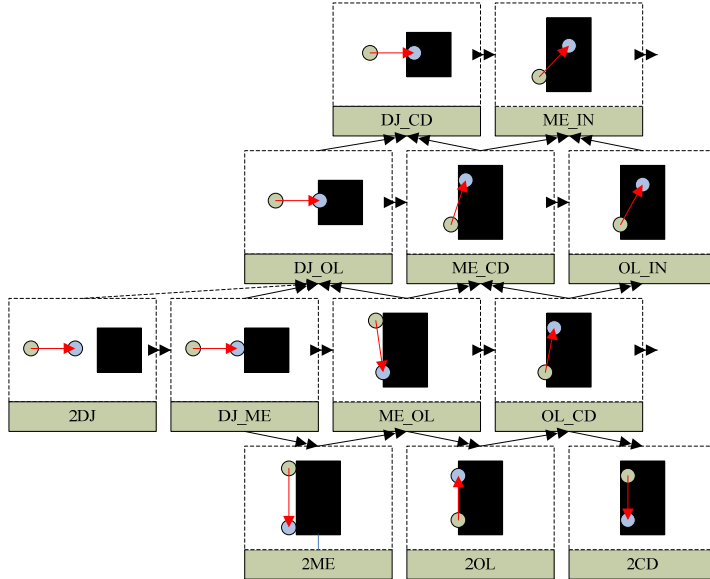


Fig. 5. Part of Proposed Motions

#### 4.2 Modeling of Directional Relation Motions

Some case gets a directional relation such as table 4 and fig. 6. They represent a different motion following final position. In this paper, we just consider 4 directions such as fig. 4.

Table 4. Topological Relation including Directional Relation

No.	Initial		Final		Motion
	Topology	position	Topology	position	
1	disjoint	1	disjoint	1	return
2	disjoint	1	disjoint	2	go through
3	disjoint	1	disjoint	3	turn left
4	disjoint	1	disjoint	4	turn right

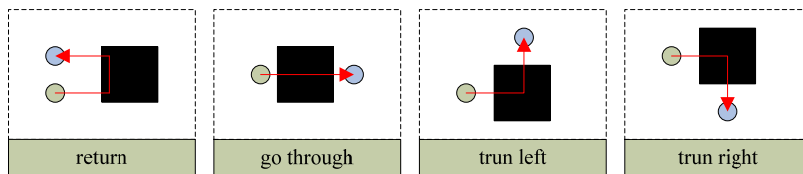


Fig. 6. Motions including Directional Relation

## 5 Similarity Measure between Motions

In this section, we introduce the method of similarity measure non-directional relation and directional relation. Similarity Measure Non-Directional Relation model is using our proposed section 4 and Directional Relation model is using TSR(Tangent Space Representation).

### 5.1 Similarity Measure Non-Directional Relation Model

We proposed spatio-temporal relation in previous section. This relation is continuously moving of object following temporal relation. In fig 7, similarity (distance) is 9 from DJ1\_DJ1 (number is position of table 1 and 2) to DJ2\_DJ2 and it is 10 from DJ1\_DJ1 to go through(directional relation). We can get a similarity between moving objects. For example similarity is 5 from CD1\_IN to CD2\_OL2. Another case, similarity is 3 from DJ1-DJ1 to ME1\_IN.

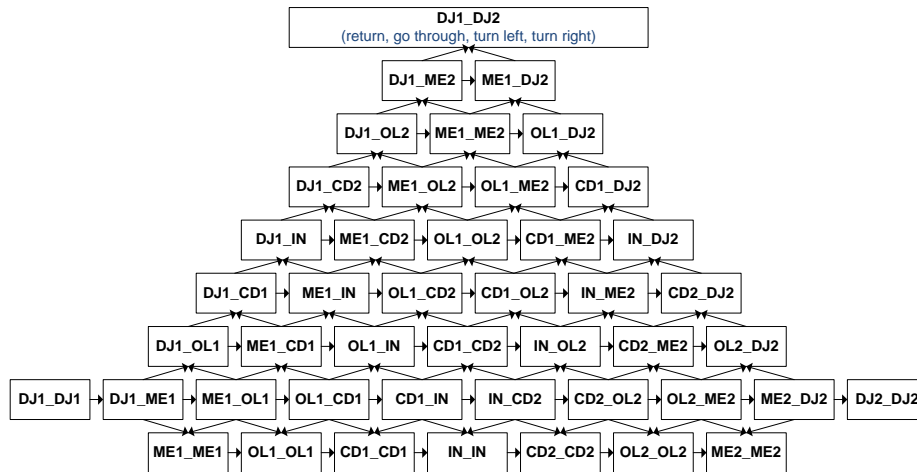
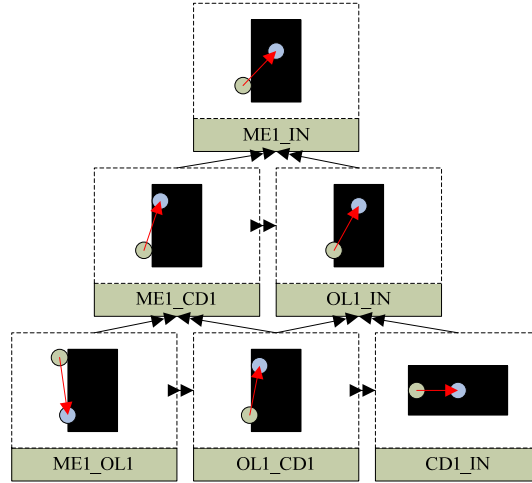


Fig. 7. Proposed Spatio-temporal Relation

For the similarity, we are using the LCM (Least Common Multiple). For example, let's get a similarity between ME1\_IN and DJ1\_DJ1. ME1\_IN can be divided ME1\_CD1 and OL1\_IN. Also ME1\_CD1 is can be divided ME1\_OL1 and OL1\_CD1 such as fig. 8. OL1\_IN is a union between OL1\_CD1 and CD1\_IN. DJ1\_DJ1 is a basic element and cannot divide. Using this, we can get a 3 distance – 2, 3 and 4 (between ME1\_OL1 and DJ1\_DJ1, between OL1\_CD1 and DJ1\_DJ1, between CD1\_IN and DJ1\_DJ1). Therefore similarity is  $1/3(2+3+4) = 3$ .





**Fig. 8.** Proposed Spatio-temporal Relation

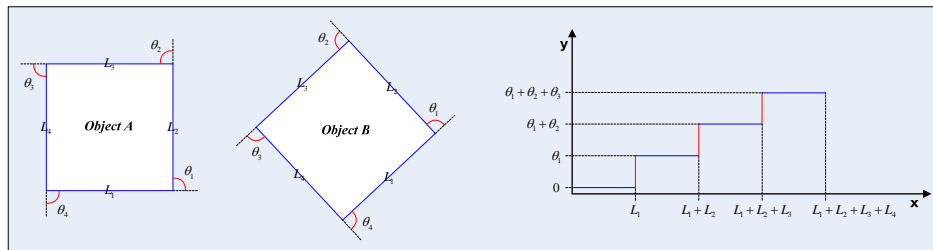
$$ME1\_IN = ME1\_OL1 + OLI\_CD1 + CD1\_IN$$

$$DJ1\_DJ1 = DJ1\_DJ1 + DJ1\_DJ1 + DJ1\_DJ1$$

$$\text{Similarity} = 1/3 ( 2 + 3 + 4 ) = 3$$

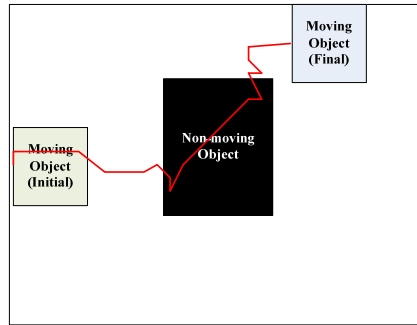
## 5.2 Similarity Measure Directional Relation Model

TSR(Tangent Space Representation)[9][10] is a kind of method to measure similarity about a object's shape in two or three dimensions. Actually, polygonal shapes consist of lines and the line also contains points, so we used TSR for measuring similarity between trajectories. In the figure 9, *Object A* and *Object B* is a same shape. The right graph is TSR about *Object A* and *Object B*. The abscissa is sum of distance about lines and the ordinate is sum of angle between lines. In this graph, there isn't difference about area so they are same shape. In this way, we applied to measure similarity between trajectories using TSR.



**Fig. 9.** Similarity Measure Methods of Object Shape

In the figure 10, the left part shows sample motions about ‘turn\_left’. The Result (table 5) is that they are the same semantic so the right figure shows the same graph about a group of ‘turn\_left’. We get the value after measuring similarity between trajectories. In this study, if two trajectories are the same, the similarity between those is 0. In other words, similarity is in inverse proportion to measured value. The sample trajectory is the most similar with ‘turn left’ in table 5.



**Fig. 10.** Sample trajectory of moving object

**Table 5.** Similarity Measure using sample trajectory

	go_through	<b>turn_left</b>	return
Similarity	51.26031	<b>30.4637</b>	97.0059

## 6 Experiment

We experiment on semantic-based trajectory retrieval with the proposed method. For the experiment, we make 100 coordinate set (non-moving object’s coordinate set and moving object’s trajectory) using simulator. The result is 97% (3% is simulator error) because is not a real video data. This paper is a just proposed Extended Spatio-temporal relation so we did use a real video data because problem of object detection and trajectory extraction. Therefore the result doesn’t have a any meaning. It just checks our model. In previous work[3], we got a 92.857% using previous model + TSR and when we use proposed model, the precision rate increase to almost 100%.

Difference between previous work and this work as following :

- Number of models (before: 6, now: 51) – except directional relation and size of moving object. Therefore, previous model cannot represent all of case but this work considered the number of all cases.
- Max distance is just 4 in previous work and it is 10 in this work so we can get similarity of various models in non-directional relation.
- This research is inserted initial and final position so it’s better than before work for the retrieval of motions such as intrusion detection using CCTV.

## 7. Conclusion and Future Works

In this paper, we proposed new spatio-temporal relation. Previous work, we cannot representation all of models but we considered the number of all cases in this study. Increasing number of models, we limit a mapping between motions and motion verbs so we define motion verb about the 14 basic model and others are representation using topological relation and position. We be sure when applying intrusion detection system, get a good result therefore next study we will apply this using real video data and CCTV.

**Acknowledgments.** This work was supported by the Korea Science and Engineering Foundation (KOSEF) grant funded by the Korea government (MEST) (No. 2009-0064749) and by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MEST) (No. 2010-0011656)

## References

1. M.Egenhofer and R.Franzosa, "Point-set topological spatial relations", International Journal of Geographical Information Systems, vol. 5, issue 2, pp.161-174, 1991.
2. Martin Erwig and Markus Schneider, "Spatio-Temporal Predicates", IEEE TKDE, Vol. 14, No. 4, pp 881-901, 2002
3. Miyoung Cho et al, "Trajectory Annotation and Retrieval based on Semantics", pp 251-264, AMR 2007, LNCS 4918, 2007
4. Anthony J.T. Lee, Han-Pang Chiu and Ping Yu, "Similarity Retrieval of Videos by using 3D C-String knowledge Representation". Journal of Visual Communication & Image Representation, pp 749-773, 2005
5. J.W. Chang, Y.j. Kim and K.J. Chang, "A Spatial Match Representation Scheme for Indexing and Querying in Iconic Image Databases", ACM International Conference on Information and Knowledge Management, pp.169-176, Nov. 1997.
6. Jae-Woo Chang and Yeon-Jung Kim, "Spatial-Match Iconic Image Retrieval with Ranking in Multimedia Databases", Proceedings of Advances in Web-Age Information Management: Second International Conference, Jul. 2001.
7. W. Ren, M Singh and S Singh, "Image Retrieval using Spatial Context", 9th International Workshop on Systems, Signals and Image Processing, November, 2002
8. Pei-Yi Chen, Arbee L.P. Chen, "Video Retrieval Based on Video Motion Tracks of Moving Objects", Proceedings of SPIE Volume 5307, pp.550-558, 2003.
9. Sunkyong Baek, Myungwon Hwang, Miyoung Cho, Chang Choi, and Pankoo Kim, "Object Retrieval by Query with Sensibility based on the Kansei-Vocabulary Scale," Computer Vision in Human-Computer Interaction, The Proceedings of the ECCV2006 Workshop on HCI, LNCS 3979, pp.109-119, 2006.
10. Myungwon Hwang, Sunkyong Baek, Hyunjang Kong, Juhyun Shin, Wonpil Kim, Soohyung Kim and Pankoo Kim, "Adaptive-Tangent Space Representation for Image Retrieval based on Kansei", LNAI 4293, pp.828-837, MICAI 2006.
11. Miyoung Cho et al, "Comparison between Motion Verbs using Similarity Measure for the Semantic Representation of Moving Object", LNCS 4071, pp 282-290, CIVR2006
12. Zaher A. Aghbari, "Studies on Modeling and Querying Video Databases", degree of Doctorate of Philosophy, Kyushu University, 2001

13. S.Y. Lee and F.J. Hsu, "Spatial reasoning and similarity retrieval of images using 2D C-String knowledge representation", *Pattern Recognition*, vol. 25, no. 3, pp. 305-318, 1992.
14. John Z. Li, M. Tamer Ozsu, Duane Szafron, "Modeling of Moving Objects in a Video Database", In *Proceedings of the International Conference on Multimedia Computing and Systems*, pp.336-343, 1997.
15. Wolter Pieters, "Representing Humans in System Security Models: An Actor-Network Approach", *Journal of Wireless Mobile Networks, Ubiquitous Computing, and Dependable Applications*, vol. 2, no. 1, pp 75-92, 2011
16. Jingbo SUN, Yue WANG, Hongbo SI, Jian YUAN and Xiuming SHAN, "Aggregate Human Mobility Modeling Using Principal Component Analysis", *Journal of Wireless Mobile Networks, Ubiquitous Computing, and Dependable Applications*, vol. q, no. 2/3, pp 83-95, 2010