

APPLYING AUGMENTED ORIENTATION SPATIAL SIMILARITY RETRIEVAL IN PICTORIAL DATABASE

Xiao Ming Zhou

Sybase Asia Development Center

xzhou@sybase.com

Chuan Heng Ang, Tok Wang Ling

School of Computing

National University of Singapore

angch@comp.nus.edu.sg, lingtw@comp.nus.edu.sg

Abstract Similarity-based retrieval of images is an important task in many image database applications. A major class of users' requests requires retrieving those images in the database that are spatially similar to the query image. We propose an approach for computing the orientation spatial similarity between two symbolic images in this paper. The proposed approach is not only rotation invariant, but also captures the relative distance and orientation range between objects.

Keywords: Augmented orientation Similarity retrieval Pictorial database

1. INTRODUCTION

In many applications (Huang, 1996; Lee, 1989; Papadias, 1994), images comprise the vast majority of acquired and processed data. An image of a picture is often stored and represented in two forms, the original full-sized image and its symbolic form (Chang, 1987). The symbolic version of the image is an object-based image file that consists of icons that represent the objects in the picture and features extracted from the original picture. So,

when a user queries a pictorial database, the retrieval procedure will match the query against the symbolic database first. If the features of some symbolic database pictures satisfy the query selection criteria, the real images are then retrieved and displayed.

The effectiveness of this kind of image database retrieval depends on the correctness and types of image feature representation. Basically, there are two kinds of features of an image: visual features (Ooi, 1998) and spatial relationship features (Nabil, 1996). Visual features, such as texture, shape, and color, could potentially be used as a basis for coarse-grained image similarity retrieval (e.g., *Find images which are predominantly orange*). If we extracted visual features for individual objects within pictures, we could ask more fine-grained queries (e.g., *Find images containing a red ball*). However, it is not until we introduce spatial relationship (Ang, 1998; Nabil, 1995; Sistla, 1995) into queries that we can ask very precise queries such as "*Find images containing a red ball on the top of a yellow cube.*" While a general-purpose image query will most likely involve both visual and spatial relationship feature, in this paper we concentrate only on the later.

So far, there are two kinds of representation for spatial relationship features:

(1). *Topological relationships* which are invariant under topological transformation of the reference objects (Egenhofer, 1991).

(2). *Orientation relationships* which concern partial and total orientation relationships among objects. Ang(1998) and Peuquet(1987) have developed several different models.

Topological relationship has been intensively studied in these years. In particular, it was applied in GIS due to its invariance under transformation. Egenhofer(1991) proposed a model called 4-Intersection to formally define the topological relationships which stay invariant under topological transformation such as translation, rotation, and scaling. According to this model, each object is represented in 2D space as a point set which has an interior and a boundary. Eight meaningful relationships can be defined in this model, namely disjoint(dt), touch(to), overlap(ov), covers(co), inside(in), coveredby(cb), contains(ct), and equal(eq). He has also subsequently extended his 4-intersection model to 9 intersections. This method has been popularly accepted for the representation of topological relationships.

However, it is hard to reach a consensus on the representation of orientation relationships as the variance of transformation was thought to be the intrinsic attribute of orientation. In this paper, we will concentrate on orientation relationship and propose an approach that may represent and retrieve orientation relationship invariant under transformation(Petraglia, 1993; Petraglia, 1994).

The remainder of this paper is structured as follows: Section 2 discusses orientation relationship. Section 3 proposes the **Augmented Orientation Spatial Relationship** (AOSR) representation. Section 4 discusses AOSR similarity retrieval. Section 5 describes our experiments for the proposed similarity approach. The paper is concluded in Section 6.

2. ORIENTATION SPATIAL RELATIONSHIP

Orientation relationships describe where objects are placed relative to one another. Three elements are needed to establish an orientation: a primary object, a reference object and a frame of reference. In pictorial database, orientation relationship information is more useful than topological relationships. To generate a description of a picture based on the objects' relationships automatically, a possible assumption is that all objects in the picture are topologically disjoint. This is practical because non-disjoint objects are hard to recognise.

As we know, orientation must be given with respect to a reference frame, i.e., the orientation that determines the direction in which the primary object is located in relation to the reference object. In a picture in which "*The ball is in front of the car*", the reference frame is based on the car and its front is clearly defined. There are three types of reference frames in the use of projective spatial prepositions in natural language: intrinsic, extrinsic, and deictic(Hernandez, 1994):

(1). Intrinsic orientation: The orientation is given by some inherent property of the reference object (e.g. The ball is in front with respect to the car's front). Criteria for determining the intrinsic orientation of objects and places are among others: the characteristic direction of motion or use, the side containing perceptual apparatus, the side characteristically oriented towards the observer, and the symmetry of object.

(2). Extrinsic orientation: The orientation on the reference object is imposed by external factors. Relevant factors are the accessibility of the reference object, its motion (or that of the observer), other objects in its vicinity or the earth gravitation (e.g., The ball is in front with respect to the actual direction of the motion of the car. Thus if the car is moving backwards, that direction is considered "front").

(3). Deictic Orientation: The orientation is imposed by the point of view from which the reference object is seen (by an observer within the scene or from the speaker's point of view).

Normally, the reference frame with respect to which the orientation is determined can be the combination of the above three different types. Among these three reference frames, the intrinsic orientation is the most

natural framework of reference for the relative orientations among objects. For example, the intrinsic front of a house is usually determined by the main entrance of the house, and it is the part of the scene that can be seen by looking outward from the main entrance. By using the intrinsic orientation, we don't need to worry about the rotation of the pictures. However, not all objects have intrinsic orientations. For example, it is hard to define an intrinsic front of a round table or a football. For this kind of objects, their self-orientations are not very important to other objects, and hence an extrinsic orientation can be applied to them. For the ease of discussion, we define these two kinds of objects as follows:

Definition 1: An object with an intrinsic orientation is a regular object. An object without an intrinsic orientation is called a non-regular object.

In the following discussion, we will restrict the images to contain only regular objects first. We will touch on how to add non-regular objects into the symbolic images later on. For simplicity, all objects considered are without holes and segments.

3. AUGMENTED ORIENTATION SPATIAL RELATIONSHIP (AOSR) REPRESENTATION

Usually orientation categorization is based on the relative positions of the centroids of objects. However, for objects with extension, this kind of orientation relationship can't reflect the meaningful information between objects. For example, in Fig.1(a), if we simply say object B is at the northeast of object A according the centroids of A and B, then there will be no difference between Fig.1 (b), (c) and (a). However, from a human's viewpoint, they are different. In Fig.1(a), if the size of object B changes, the orientation range of object B relative to object A will change too. In Fig.1(b), the relative distance between object A and B is different from that in Fig.1(a). The intrinsic orientations of object B in Fig.1(a) and Fig.1(c) respectively are not the same. Therefore, in the following of this paper, we propose an **Augmented Orientation Spatial Relationship (AOSR)** representation to capture this information.

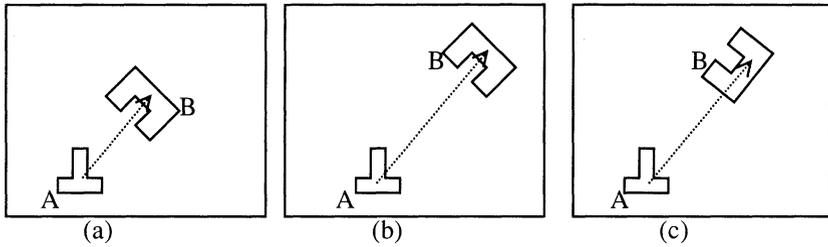


Figure 1. Example pictures

3.1 DEFINITION OF AOSR

The augmented orientation spatial relationship representation (AOSR) between a reference object and a primary object is defined as follows:

Definition 2: The **Augmented Orientation Spatial Relationship (AOSR)** between a reference object O_r and a primary object O_p is a tuple of α and β , the angles of the begin-bound and the end-bound respectively with $\beta > \alpha$. The begin-bound is the first tangent line which is about to enter object O_p when rotating anti-clockwise a half line originated from the centroid of object O_r . The end-bound is the first tangent line which is about to leave object O_p when rotating anti-clockwise a half line originated from the centroid of object O_r . The starting direction of the rotating half line is the **intrinsic front** of object O_r . The relationships between O_r and itself is defined to be NULL.

The illustration of the above definition is shown in Fig. 2. With this definition, we may capture the orientation spatial relationships of objects with extension more precisely because the begin-bound angle and the end-bound angle denote the extension of the primary object in the polar coordinate system of the reference object. In the meantime, the relative distance information of the primary object and the reference object can also be reflected by the begin-bound angle and the end-bound angle combined with the intrinsic orientations of both objects. In Fig.3, when we move object O_p away or toward object O_r along the line between the two centroids of O_r and O_p , the begin-bound angle and the end-bound-angle will change. The relative distance of O_r and O_p can be reflected by the difference $\beta - \alpha$ if both O_r and O_p 's intrinsic fronts remain unchanged.

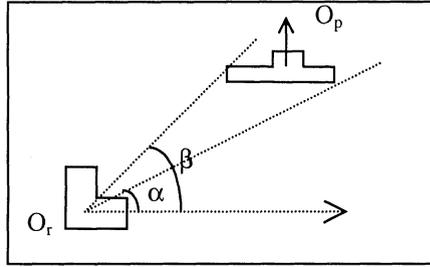


Figure 2. Definition of α and β for O_r and O_p (The arrows represent the intrinsic fronts of objects)

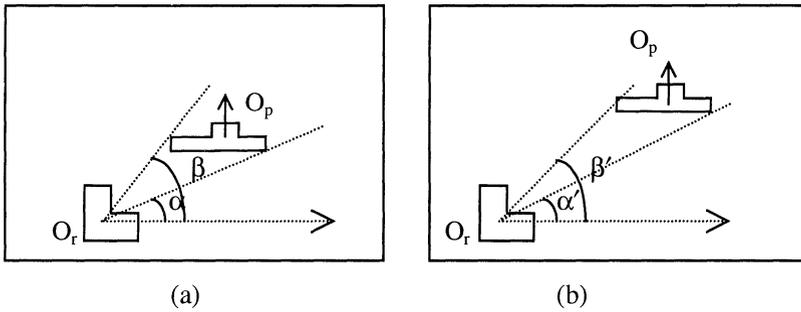


Figure 3. An example (the arrow represents intrinsic front of O_p)

However, usually the begin-bound angle and the end-bound angle will change too when the orientation of object O_p is changed. An example is shown in Fig.4. This will cause ambiguity because the change in either the relative distance or the orientation may result in the same AOSR for a reference object O_r and a primary object O_p . The problem can be solved if we use **intrinsic orientation**. For example, with the use of **intrinsic orientation** of the reference object as the reference frame, even if the begin-bound angle and the end-bound angle remain the same, the change in the intrinsic orientation of the primary object O_p in Fig.4 can still be detected. This will cause the orientation spatial relationship to change when O_p is used as reference object. Hence it is very important that the intrinsic orientation is used in the AOSR to measure angles.

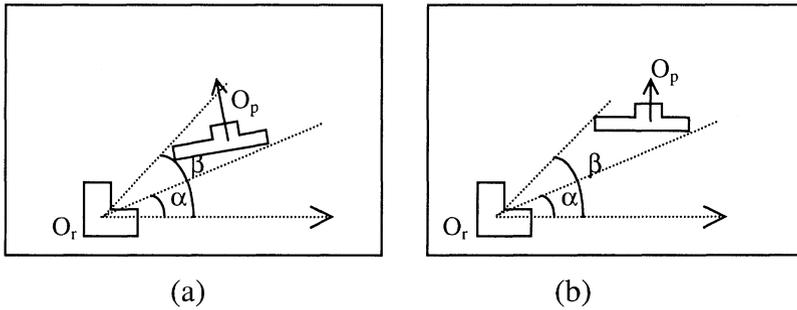


Figure 4. An example (the arrow represents intrinsic front of O_p)

In the above, what we have defined is the AOSR between a primary object and a reference object. Given a picture with n different objects, each of the n objects can be chosen as a reference object with the rest as primary objects. Therefore, we can define the AOSR of a picture as follows:

Definition 3: For a picture with n different objects, its **AOSR** representation is an $n \times n$ matrix where each row is the AOSR between the row index object as the reference object and each of the n objects as the primary objects. The row index objects and the column index objects are lexicographically sorted.

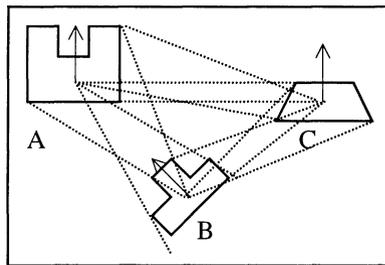


Figure 5. An example picture P (all arrows represent intrinsic fronts of objects)

In the above definition, it is obvious that the matrix is built based on all the objects involved since each row of the matrix is the set of the orientation spatial relationships of all objects using the row index object as the reference object. As an example, the AOSR of picture P in Fig.5 is as follow (For the ease of illustration, all angles are of integer values):

	A	B	C
A	NULL	(210, 240)	(255, 270)
B	(-15, 15)	NULL	(240, 270)
C	(45, 95)	(210, 235)	NULL

Note that the matrix entry $P(B, A) = (-15, 15)$ instead of $(345, 15)$ as β must be bigger than α as stated in definition 2.

3.2 HANDING NO-REGULAR OBJECT

In some cases, there is a mixture of regular and non-regular objects. To construct each entry of the AOSR matrix, we need to handle the following two cases:

1. One of the objects concerned is non-regular.

In this case, we use the regular object's intrinsic front as the non-regular object's intrinsic front.

2. Both objects are non-regular objects.

In this case, α is always 0 and β is the subtended angle of the primary object with respect to the centroid of reference object.

It is very straightforward to use only existing regular objects' intrinsic orientation as the reference frame in the AOSR matrix since non-regular object's self-rotation won't affect the spatial relationship between objects usually. However, for pictures containing only non-regular objects, our AOSR is not applicable at this stage. Therefore, we assume all pictures discussed in this paper contain at least one regular object. A special case is that there is only one regular object in a picture. In this case, the intrinsic orientation of the regular object is the only reference orientation frame in the picture.

4. SIMILARITY RETRIEVAL

There are two kinds of picture retrieval: exact matching and similarity retrieval, with the former a special case of the latter. In similarity retrieval, we are to retrieve from the pictorial database those pictures that are similar to the query picture given by a user. To do this, we must design a real-valued function that can measure the similarity of every pair of pictures. For the ease of discussion, we assume that all objects involved are distinct in each picture. We ignore the situations in which there are multiple instances of the same type of object (such as 2 cats etc.) within a picture. It has been proved (Tucci, 1991) that the similarity retrieval problem is NP-hard when multiplicity of objects is allowed in picture matching.

Similarity retrieval is a fuzzy issue. Usually different similarity functions are used for different comparison objectives. Since our AOSR matrix representation is based on object, our retrieval procedure will also be based on object. In fact, this follows human's retrieval procedure in the real life. To compare two pictures, we normally compare them by portions using different object as reference object in each comparison. Therefore, our object-based representation fits well with the object-based retrieval.

Without loss of generality, consider two symbolic pictures Q and P with Q the query image and P the database image. To define a similarity based on object O_1 in Q, the first step is to check whether P contains O_1 . The second step, the corresponding rows and columns indexed by O_1 in the matrices of P and Q are used to compute the similarity. Following Gudivada's(1995) assumption, the maximum similarity is set to 100. If there are n objects in Q, each of the n-1 tuples involving other objects' orientations with respect to O_1 contributes a value of $100/(n-1)$ toward the similarity based on object O_1 . Suppose O_2 is another object that appears in both P and Q. $P(O_1, O_2)$ denotes the item indexed by O_1 and O_2 in picture P's AOSR matrix. If we have:

$$P(O_1, O_2) = (\alpha_{12}^P, \beta_{12}^P)$$

$$P(O_2, O_1) = (\alpha_{21}^P, \beta_{21}^P)$$

$$Q(O_1, O_2) = (\alpha_{12}^Q, \beta_{12}^Q)$$

$$Q(O_2, O_1) = (\alpha_{21}^Q, \beta_{21}^Q)$$

then the four pairs of angles α_{12}^P and α_{12}^Q , β_{12}^P and β_{12}^Q , α_{21}^P and α_{21}^Q , and β_{21}^P and β_{21}^Q will be used to measure the similarity based on O_1 and O_2 . When the angular difference between a pair of angles increases from 0 to 180, the similarity contributed by this pair of angles will decrease from the maximum to the minimum. However, when the angle difference between one pair of angle increases from 180 to 360, the similarity contributed by this pair of angles will increase from the minimum to the maximum. Hence we choose the cosine value of the difference for the pair of angle plus 1 to ensure the value to be nonnegative. The contribution factor from O_2 to the similarity based on O_1 is defined as

$$S_{12} = 100*((\cos(\alpha_{12}^P - \alpha_{12}^Q) + 1) + (\cos(\beta_{12}^P - \beta_{12}^Q) + 1) + (\cos(\alpha_{21}^P - \alpha_{21}^Q) + 1) + (\cos(\beta_{21}^P - \beta_{21}^Q) + 1))/9(n-1) \quad (1)$$

In formula (1), when $\alpha_{12}^P = \alpha_{12}^Q$, $\beta_{12}^P = \beta_{12}^Q$, $\alpha_{21}^P = \alpha_{21}^Q$, $\beta_{21}^P = \beta_{21}^Q$, the contribution factor is $100/(n-1)$, where n is the number of objects in the query picture. This is the maximum similarity contributed by one object. When the difference between the corresponding angles in P and Q is bigger, the similarity contribution is smaller. It should be noted that the minimum similarity contribution is not 0 but $100/9(n-1)$ when all the four cosines' values are -1.0. This situation happens when all corresponding angles are

different 180 degree. However, if either O_1 or O_2 does not appear in P, the contribution factor from O_2 is defined as 0. This means that as long as O_2 appears in P too, the contribution factor from O_2 to S_{12} must be bigger than 0, which is a very important characteristic of the similarity measure.

Now we may replace O_2 by any other object appearing in Q. Then S_1 , the total similarity based on object O_1 , is the sum of all contributions from all other $n-1$ objects in Q.

$$S_1 = \sum S_{1j} \quad (j= 2 \dots n, \text{ where } n \text{ is the number of objects in Q}) \quad (2)$$

The maximum S_1 is 100 and the minimum S_1 is 0 when O_1 is missing from P or all other $n-1$ objects are missing. The value of S_1 indicates how spatially similar is P to Q based on O_1 .

Suppose the AOSR matrices for Q and P are M_Q and M_P respectively. We define a similarity function SV:

$$SV(M_Q, M_P) = (S_1, S_2, \dots, S_n) \quad \text{where } n \text{ is the number of objects in Q} \quad (3)$$

The SV() uses M_Q and M_P as inputs to compute the similarity degrees between Q and P based on each of the objects in M_Q .

How to compute the spatial similarity between pictures by using the n similarity values produced by SV() is application dependent. Here are two possible approaches:

1. Selecting some **dominant objects** in the query image to compare similarity. For example, in the query to *find a house, in the front of which there is a temple and a swimming pool both facing this house*, the house is the dominant object. The relationship between the temple and the swimming pool is not mentioned and is assumed to be unimportant. In this case, not all values from the SV() are useful. In a real application, it may not be necessary to compute all S_1 to S_n . This approach can trim the answer set of similar pictures very effectively. It is especially useful for similarity retrievals involving only a small number of dominant objects
2. Comparing the similarity based on the sum of all S_1 to S_n . This approach is widely used for the image retrieval without dominant objects. The result is a small set of candidate images ranked according to their similarity degrees.

We implemented both retrieval approaches in the following experiments.

5. EXPERIMENTS

Similarity retrieval is always fuzzy and subjective. It is hard to find a widely accepted benchmark for similarity retrieval experiments. In our experiment, we try to construct our experimental database to include many different pictures organised into several groups.

There are 647 pictures in our database. Each image contains 10 to 15 objects. The pictures are divided into 5 groups according to the number of objects' changes (include moving and rotating etc.) with respective to the query picture. There are 20, 45, 120, 210, and 252 pictures in group 1, 2, 3, 4, and 5 respectively. A picture is in group i when there are i objects changing their positions with respect to the query image. Those pictures with five or more objects changing their positions belong to group 5.

For the first experiment, we randomly choose one unchanged object as dominant object, and calculate for each group the average percentage of pictures that the similarity degree is more than 90, 93, 95, 98 and 99 respectively. The results are shown in Table 1.

For the second experiment, we calculate all S_1 to S_n and sum the total to get the top 10, the top 20, the top 30, and the top 40 pictures with higher similarity and compute the percentage distribution of images that have been retrieved from each group. Table 2 shows the results of the experiment.

Table 1. Experiment 1 results

Similarity	The percentage of pictures retrieved (%)				
	Group 1	Group 2	Group 3	Group 4	Group 5
>90	100.00	78.67	62.33	48.48	37.06
>93	100.00	65.33	45.83	28.95	18.41
>95	100.00	55.56	34.00	19.62	9.52
>98	100.00	38.67	14.67	4.67	1.35
>99	100.00	25.78	5.33	0.85	0.08

Table 2. Experiment 2 results

Ranks	The distribution percentage among groups				
	Group 1	Group 2	Group 3	Group 4	Group 5
10	100.00	0.00	0.00	0.00	0.00
20	85.00	15.00	0.00	0.00	0.00
30	66.67	26.67	6.67	0.00	0.00
40	50.00	32.50	15.00	2.50	0.00

From Table 1 and 2, we find that the percentage of pictures retrieved from groups with fewer object changes is always higher. When we relax the similarity degree requirement for experiment 1 or rank more pictures for experiment 2, there are more and more pictures being retrieved from groups

other than the first group. This shows that even when many objects have shifted their positions, as long as their extent of changes remain small, they will be retrieved as pictures similar to the query pictures, yet they may be rejected when other algorithms such as (Nabil, 1996) are being used.

There are two factors to be considered when we determine whether a picture is similar to another. One is the number of objects that have changed their spatial relationships, and the other is the extent of these changes. Our experiment shows clearly that the number of objects that have changed their positions is not the only factor that affects the spatial similarity. Although in practice, a picture with many objects changed their positions will be more likely to be rejected in similarity retrieval, one can argue that it should be accepted if the extent of changes is small. Our experiment results show that these factors complement each other.

6. CONCLUSIONS

In summary, we have proposed in this paper a similarity retrieval approach for the augmented orientation spatial relationship (AOSR) representation. Comparing to existing systems, the proposed approach is not only rotation invariant, but also captures the relative distance and orientation range between objects. It overcomes the ambiguity problems that exist in other orientation representations, and is more flexible and applicable.

REFERENCES

- Ang, C. H., T.W. Ling, and X.M. Zhou, 1998. Qualitative spatial relationships representation IO&T and its retrieval. *9th International Conference, DEXA'98, Vienna, Austria, August 1998. Lecture Notes in Computer Science* 1460, 270-279
- Chang, S.K., Q.Y. Shi, and C.W. Yan, 1987. Iconic indexing by 2D strings. *IEEE Transaction on Pattern Recognition and Machine Intelligence*, Vol. 9. No.3, 413-428
- Egenhofer, M.J. 1991. Point-set topological spatial relations. *International Journal of Geographical Information Systems*. Vol. 5. No.2, 161-174
- Gudivada, V., V. Raghavan, 1995. Design and evaluation of algorithm for image retrieval by spatial similarity, *ACM Transaction on Information Systems*, vol. 13, no.1, 115-144
- Hernandez, D. 1994. Qualitative representation of spatial knowledge. *Lecture Notes in Artificial Intelligence* no. 804, Springer -Verlag
- Huang, P.W., Y.R. Jean, 1996. Spatial reasoning and similarity retrieval for image database systems based on RS-string, *Pattern Recognition*, Vol. 29, No. 12, 2103-2114
- Lee, S.Y., M.K. Shan, and W.P. Yang, 1989. Similarity retrieval of iconic image databases. *Pattern Recognition*, Vol. 22, No. 6, 675-682
- Nabil, M., J. Sheperd, and A.H.H. Ngu, 1995. 2D projection interval relationships: A symbolic representation of spatial relationships. *Advances in Spatial Databases: Fourth Int'l Symp., SSD'95, Lecture Notes in Computer Science* no.951, 292-309, Springer-Verlag

- Nabil, M. A.H.H. Ngu, and J. Sheperd, 1996. Picture similarity retrieval using 2D projection interval representation. *IEEE Transactions on Knowledge and Data Engineering*. Vol.8, No.4, 533-539
- Ooi, B.C., K.L. Tan, T.S. Chua, and W. Hsu, 1998. Fast Image Retrieval Using Color-Spatial Information, *The VLDB Journal*, Springer-Verlag, Vol. 7, No. 2, 115-128
- Papadias, D., T. Sellis, 1994. A pictorial language for the retrieval of spatial relations from the image databases. *Proc. Of the 6th International Symposium on Spatial Data Handling*. Taylor Francis
- Petraglia, G., M. Sebillio, M. Tucci, and G.Tortora, 1993. Towards normalized iconic indexing. *Proc. of the 1993 IEEE Symposium on Visual Language*, Bergen, Norway, 392-394
- Petraglia, G., M. Sebillio, M. Tucci, G. Tortora, 1994. Rotation invariant iconic indexing for image database retrieval. *Progress in Image Analysis and Processing III* (edited by S. Impedovo), World Scientific, 271-278
- Peuquet, D.J., Z. Ci-Xiang, 1987. An algorithm to determine the directional relationship between arbitrary-shaped polygons in the plane. *Pattern Recognition*, Vol. 20. No. 1, 65-74
- Sistla, A.P., C. Yu, C. Liu, and K. Liu, 1995. Similarity based retrieval of pictures using indices on spatial relationships. Proc of the 21th VLDB Conference, Zurich, Switzerland
- Tucci, M., G. Costagliola, and S.K. Chang, 1991. A remark on NP-completeness of picture matching, *Information Processing Letters*.

BIOGRAPHIES

Xiao Ming Zhou is with Sybase Asia Development Center currently. Before joining Sybase, he was an academic staff in School of Computing, National University of Singapore. He received his Bachelor's degree from Tongji University (China), and M.Sc. from National University of Singapore. He is also a Ph.D candidate.

Chuan Heng Ang is a senior lecture in School of Computing, National University of Singapore. Dr. Ang received his Bachelor's degree and M.Sc from Nanyang University (Singapore), and Ph.D. from University of Maryland.

Tok Wang Ling is a professor in School of Computing, National University of Singapore. Dr Ling is also a Vice Dean of the School. He received his Ph.D. and M.Math., both in Computer Science, from Waterloo University (Canada) and B.Sc.(First Class Hons) in Mathematics from Nanyang University (Singapore).