A Multi-step Strategy for Shape Similarity Search In Kamon Image Database

Paul W.H. Kwan[1], Kazuo Toraichi[2], Keisuke Kameyama[2], Junbin Gao[3], Nobuyuki Otsu[4]

[1] School of Mathematics, Statistics and Computer Science, University of New England, Australia
[2] Graduate School of Systems and Information Engineering, University of Tsukuba, Japan
[3] School of Information Technology, Charles Sturt University, Bathurst, Australia
[4] National Institute of Advanced Industrial Science and Technology (AIST), Japan

Email: kwan@mcs.une.edu.au

Abstract

Similarity search in image databases relies on comparing the query with a set of images based on features like shape, colour, texture, and spatial locations. As the size of database grows, query processing strategies were proposed to increase performance by reducing the number of distance calculations. Most strategies are two-step, with the initial "prune" step based on a high-dimensional spatial index followed by a "refine" step performing expensive computation. They work well with metric similarity models where lower bounding distance functions exist for pruning. In this work, similarity search in a Japanese Kamon Image Database is attempted. The choice of shapes as features is deliberate because kamons are in black and white, and their meanings are conveyed by shapes. Further, a three-step "prune-filter-refine" strategy targeting models with non-metric distance functions is described. Compared to the two-step approach, this strategy achieves a further reduction in number of distance calculations needed but with close to no change in the precision figure.

Keywords: Similarity Search, Image Database, Multi-step Strategy, Non-metric Distance, Autocorrelation

1 Introduction

As the size of image database in practical applications continues to grow, it has become clear that sequential comparison might not scale to very large databases. This has led researchers to propose query processing strategies that were meant to improve performance by pruning the search space prior to making expensive distance calculations. In practice, these strategies are largely two-step, with the initial "prune" step based on a high-dimensional spatial index followed by a "refine" step employing expensive computations [1].

Most existing strategies assume a metric similarity model for which a lower bounding distance function can be defined for pruning. Recent research on robust image matching methods for appearance-based vision has however confirmed that similarity models behind human visual judgment are inherently non-metric [2]. When applying these similarity models, one has to address the problem of non-metric distance functions for which optimal lower bounds might not exist.

In this work, similarity search in a Japanese Kamon Image Database that has both cultural and commercial significances is attempted. The choice of shape as the primary feature for matching is deliberate because Japanese kamons are in black and white, and their meanings are conveyed by their shapes. Furthermore, a three-step "prune-filter-refine" query processing strategy suitable for similarity models having non-metric distance functions is described. Compared with the two-step approach, this strategy achieves further reduction in number of distance calculations but with close to no change in precision.

The rest of this paper is organized as follow. Section 2 describes the steps of the proposed multi-step query processing strategy. Section 3 presents results from experiments conducted on a commercial database of 2,000 Japanese kamon images [3]. Finally, Section 4 ends with concluding remarks.

2 The Multi-step Strategy

The proposed strategy consists of three successive steps, which are collectively called the "prune-filter-refine" (or *PFR* in short) strategy. In the "prune" step, a multi-dimensional spatial index is used to eliminate improbable matches by an adjustable range threshold. Next, the "filter" step provides further reduction via a quasi lower-bounding distance that was derived from the non-metric distance function. Finally, the "refine" stage compares the remaining candidates by a robust matching method for the final similarity ranking. For ease of explanation, Figure 1 illustrates pictorially the proposed multi-step strategy for similarity search in the Japanese Kamon Image Database.

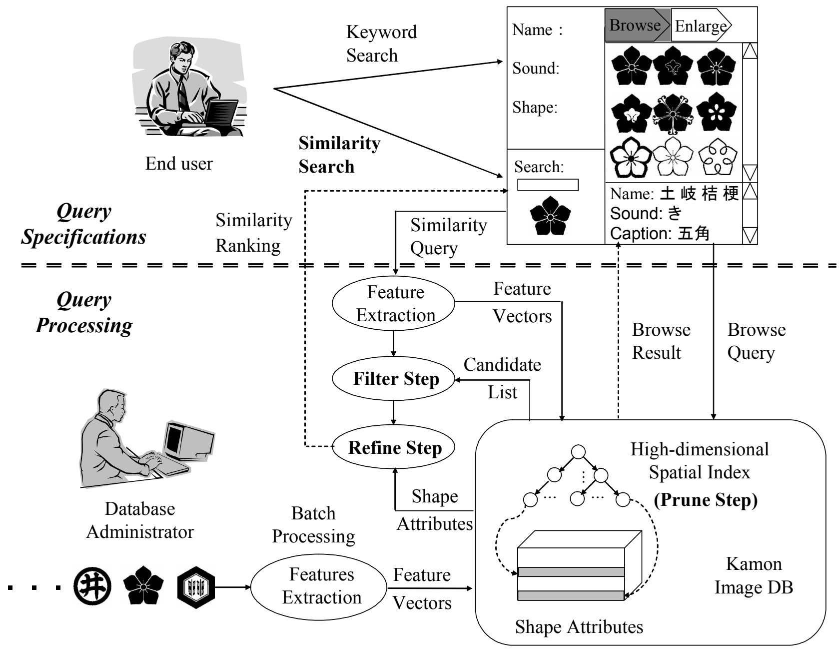


Figure 1: An overview of the proposed strategy for similarity search in the Kamon Image Database

Because both the feature vectors used in constructing the spatial index and the shape attributes used by the relaxation labeling based image matching method are output of features extraction, it will be explained before describing the steps of the proposed strategy.

2.1 Features Extraction

Kwan [4] described a contour-based image matching algorithm based on relaxation labeling. Query processing is entirely sequential, that is without using any indexing structure. In this work, Kwan [4] is applied to the filter-and-refine steps of the proposed strategy, while the initial prune step employs a multi-dimensional spatial index based on the k-d-B tree [1]. The set of numerical vectors used in constructing the spatial index is obtained by carrying out the Discrete Fourier Transform (DFT) on both the horizontal and vertical autocorrelation plots derived from an image by treating these plots as time sequence data. The algorithm used in constructing these plots is similar to Nagashima [5]. The idea is illustrated in Figure 2. Starting from a complete overlap of an image by a copy of itself, horizontal autocorrelation is measured by shifting the image one pixel at a time from left to right while calculating for each shift the degree of autocorrelations based on the number of overlapping non-background pixels. The process is repeated until there is no overlap between the image and its copy. A similar procedure is applied when the vertical autocorrelation is measured, albeit by shifting from top to bottom instead.

More formally, let x and y be variables that denote the autocorrelation lengths measured in both the horizontal and the vertical directions respectively. The horizontal and vertical autocorrelations can be expressed in terms of x and y as follow:

$$AC_h(x) = \frac{1}{l_Y}\sum_{Y=1}^{l_Y}\frac{1}{l_X - x}\sum_{X=1}^{l_X - x} f(X,Y)f(X+x,Y) \qquad (1)$$

$$AC_v(y) = \frac{1}{l_X}\sum_{X=1}^{l_X}\frac{1}{l_Y - y}\sum_{Y=1}^{l_Y - y} f(X,Y)f(X,Y+y) \qquad (2)$$

Here, $f(\cdot,\cdot)$ returns 0 for a background pixel and a 1 for a non-background pixel. Based on the two above equations, the horizontal and vertical autocorrelation plots shown in Figures 2(c) and (d) can be produced. Taking the horizontal and vertical autocorrelation plots as individual time sequences, the DFT generates for each plot a set of Fourier series coefficients in the following manner.

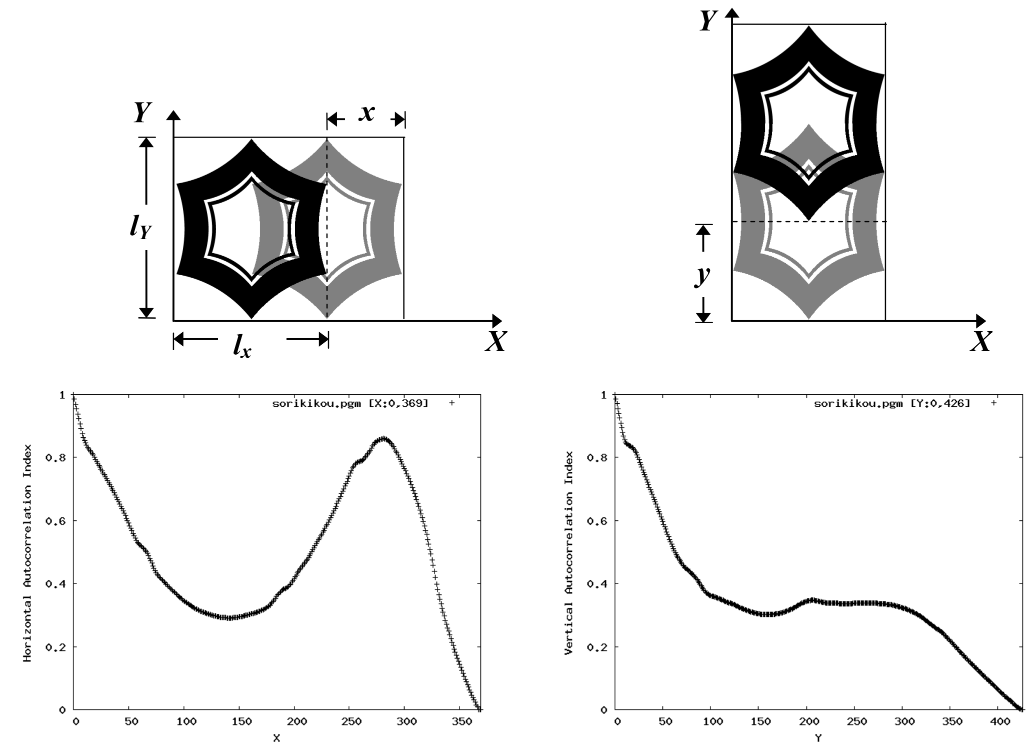
Let a time sequence $\vec{x} = [x_t]$ for $t = 0, 1, \ldots, n\text{-}1$ be a finite duration signal. The DFT of $\vec{x}$, denoted by $\vec{X} = [X_f]$, is given by:

$$X_f = \frac{1}{\sqrt{n}}\sum_{t=0}^{n-1} x_t e^{\frac{-j2\pi t f}{n}} \qquad f = 0,1,\ldots,n-1 \qquad (3)$$

Here, $j = \sqrt{-1}$ is the imaginary unit.

The inverse DFT of $\vec{X}$ returns the original time sequence by the following equation:

$$x_t = \frac{1}{\sqrt{n}}\sum_{f=0}^{n-1} X_f e^{\frac{j2\pi t f}{n}} \qquad t = 0,1,\ldots,n-1 \qquad (4)$$



a b **Figure 2:** (a) and (b) illustrate the horizontal and vertical autocorrelations of an image with itself
c d (c) and (d) are the horizontal and vertical autocorrelation plots respectively

For each image, the two sets of coefficients are concatenated to form a numerical vector that is indexed by the spatial index. In our experiments, the number of coefficients for each of horizontal and vertical autocorrelation is chosen as 15.

In this work, in addition to the numerical vectors of Fourier series coefficients, piecewise line segments that form parts of closed contours of shapes in kamon images are extracted and approximated by functions before being used as matching attributes in the filter-and-refine steps of the proposed multi-step strategy [6]. As an illustration, Figures 3(a) and (b) present a kamon image and its function approximated counterpart. Joint points detected on the contours are also shown.

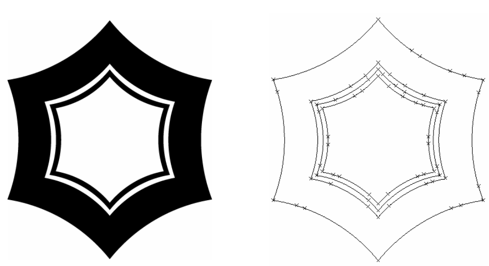


Figure 3: A Kamon Image and Its Function Approximated Counterpart by Haruki [6]

2.2 "Prune-Filter-Refine" Processing

2.2.1 The "Prune" Step

Similar to many two-step processing strategies on similarity search in the literature, the first step of the proposed strategy relies on a multi-dimensional spatial index to prune improbable database objects from further matching. The feature space consists of numerical vectors formed from a concatenation of Fourier series coefficients generated by applying DFT on the horizontal and vertical autocorrelation plots of the database images.

In this work, the k-d-B tree is chosen due to its search and I/O efficiency. When applying the k-d-B tree in indexing, the numerical vectors are assumed to be embedded in a d-dimensional *normalized* Euclidean space E^d. Let U be the universe of all such vectors. For any $O_i, O_j \in U$, their dissimilarity is defined by a distance metric $D(O_i, O_j) \in R^+$ in E^d as:

$$D(O_i, O_j) = \sqrt{(O_i^1 - O_j^1)^2 + \ldots + (O_i^d - O_j^d)^2} \qquad (5)$$

where O_i^k and O_j^k denote the attribute values of O_i and O_j in the k^{th}-dimension respectively.

Furthermore, the condition below holds for the distance function:

$$0 < D(O_i, O_j) \leq \sqrt{d}\ ,\quad \forall O_i, O_j \in U \text{ and } i \neq j \qquad (6)$$

Based on the k-d-B tree based multi-dimensional index, two types of similarity queries, namely nearest neighbour and range queries are possible. However, in the context of the proposed multi-step strategy, only range queries are relevant because the objective of the "prune" step is to reduce the search space to a much smaller candidate set for further processing by the "filter-and-refine" steps. Given $O_q \in U$ as the query, by specifying $r \in R^+$ as the range, such that $0 \leq r \leq \sqrt{d}$ holds, the candidate set (denoted C) that meets the condition below is returned:

$$C = \{O_p \mid O_p \in U \ \ and \ \ D(O_p, O_q) \leq r\} \qquad (7)$$

Lastly, as seen in Figure 1, a *pointer* is appended as the last element of each numerical vector in order to aid retrieving the corresponding shape attributes that will be used in the filter-and-refine steps.

2.2.2 The "Filter and Refine" Steps

The filter-and-refine steps of the proposed strategy are an application of Kwan [7]. Kwan [7] proposed an *approximate query processing* approach to address the performance problem of sequential matching by the method described in Kwan [4]. In simple terms, Kwan [4] matches the shape of objects between the query and each database image, and ranked them by a *heuristic* distance function. The matching algorithm was based on probabilistic relaxation labeling, in which the query image represents the set of objects while a database image the set of labels of the labeling problem.

Because the heuristic distance function is defined using probabilistic variables whose values are not known until the iterative probability updating of relaxation labeling has converged, the space of all objects is not formed until a query is entered. Further, this space is not metric in the sense of obeying the triangle inequality on distances, rendering it difficult to designate one database image as the vantage point for a possible index [1].

In Kwan [7], the core concept is a lower bounding distance function used for filtering out irrelevant database objects from expensive steps of computing actual *query distances* [8]. Whereas related work use *provable* lower bounds due to the metric nature of their similarity models, in research including ours where non-metric distance functions are used, a provable lower bound might not exist. To address this issue, a *quasi lower bounding distance* function is introduced. This function (which could be many) is defined based on the non-metric query distance function and a *confidence factor* used for scaling. In practice, it is computed at the point where the *initial state* of the relaxation labeling system is set, but prior to the iterative probability updating for actual query distance calculation has commenced. Through filtering, it is expected that a significant reduction in number of expensive query distance calculations for the "refine" step can be obtained. Further details on the filter-and-refine steps can be found in Kwan [7], and will not be repeated here.

3 Experimental Evaluation

3.1 Evaluation Criteria

A number of experiments were performed on the database of 2,000 kamon images that came with Come on Kamon Ver. 2.0 by System Product Corp. of Japan [3]. Three primary criteria for evaluating the set of experiments are:

1. The effectiveness of Fourier series coefficients derived from both the horizontal and vertical autocorrelation plots as values of numerical vectors for the multi-dimensional index.
2. The effectiveness of the combination of quasi lower bounding distance and confidence factor in reducing the number of expensive distance calculations in the filter-and-refine steps while not compromising the precision.
3. The advantage of the proposed strategy over the 2-step "prune-refine" approach in terms of reduction in the number of images matched.

3.2 Experimental Results

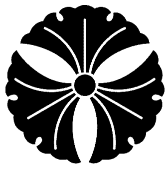
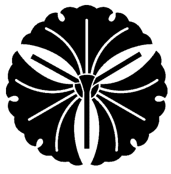
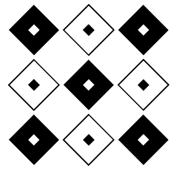
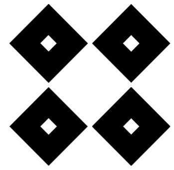
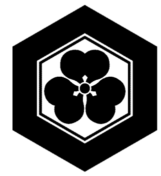
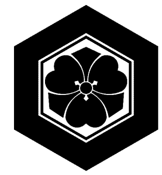
First, results shown in Tables 1 and 2 collectively addresses criterion (1) above. Table 1 are distances measured in the normalized Euclidean space E^d among the group of six kamon images shown in Table 2. In this experiment, two visually similar images from 3 of the 9 shape categories are chosen.

It is obvious from Table 1 that the distance is closer between images that are visually similar in Table 2. Though not explicitly given here, this is supported by the shape of the autocorrelation plots for these images, underlying the effectiveness of employing Fourier series coefficients in the index for pruning.

Table 1: Distance in E^d space between the set of 6 images shown in Table 2

	K1	K2	K3	K4	K5	K6
K1	0	0.64	2.37	2.40	3.09	3.31
K2	0.64	0	2.43	2.38	3.00	3.07
K3	2.37	2.43	0	1.99	3.29	3.40
K4	2.40	2.38	1.99	0	3.08	3.09
K5	3.09	3.00	3.29	3.08	0	1.08
K6	3.31	3.07	3.40	3.09	1.08	0

Table 2: Pairs of similar images from 3 of the 9 shape categories in the Japanese Kamon Image Database

Round		Quadrangle		Hexagon	
 K1	 K2	 K3	 K4	 K5	 K6

Second, to address evaluation criterion (2), retrieval result of a *k*-nearest neighbour query ($k = 20$) by the method of Kwan [4] is shown in Table 3. It will be used as the benchmark for precision in the rest of our evaluation. Note that in Table 3, D_{query} refers to the query distance computed by using the non-metric distance function of Kwan [4] while D_{index} is the distance in the indexed E^d space. While they do not exhibit full correlation, D_{index} alone could still be quite effective in approximate similarity ranking.

The second group of experiments is meant to verify how effective filtering by the quasi lower bounding distance can be achieved based on the number of database images that have to go through actual distance calculations. Two metrics are defined. The first of these is a reduction ratio defined as follow:

Re_1 = (# images actually match / # total database images) ∗ 100% (8)

The second metric is the precision defined as follow:

Pr_1 = (# correct responses in the "filtered" result / # responses returned) ∗ 100% (9)

Here, the denominator of Pr_1 is *k*, which is the number of nearest neighbours to retrieve. The precision is defined as the percentage of correct responses that are included in the "filtered" result. For each setting of the *confidence factor* [7], both reduction ratio and precision are computed. Results are summarized in Figures 4 and 5.

For $k = 20$, one could observe that the filter-and-refine steps achieved a significant reduction (more than 70%) in number of actual distance calculations while the precision is maintained (that is, $Pr_1 = 1.0$) at the point where the confidence factor is deduced automatically. Compared to this, both for $k = 10$ and $k = 5$, although the reductions are greater, the precisions suffered. Nevertheless, in applications where either the number of similar images to return is not overly small or that approximate results can be accepted, the savings in computation by the filter-and-refine steps are highly significant.

Finally, to address criterion (3), results of the previous set of experiments and the followings are combined. Although no automatic way exists for deducing an optimal value for range *r* as a function of the query that could provide maximum pruning while minimizing the number of false drops, it is still possible to heuristically deduce an approximate value for the database of our experiment through simulation. This result is summarized in Table 4. Two additional metrics are defined, denoted by Pr_2 and Re_2. Their definitions are given as follows:

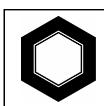
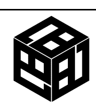
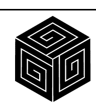
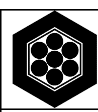
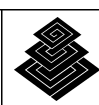
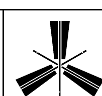
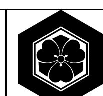
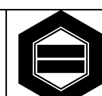
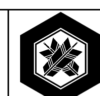
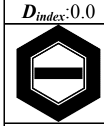
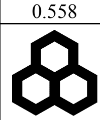
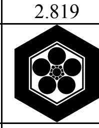
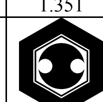
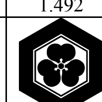
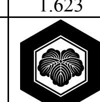
$$Re_2 = n' / n \quad (10)$$

Here, *n* and *n'* are the total sizes of the database and the number of images remained after pruning.

$$Pr_2 = k' / k \quad (11)$$

Here, *k'* is the number of answers in Table 3 that remain after pruning. Here, $0 \le Pr_1 \le 1$ holds.

Table 3: Result of a *k*-nearest neighbour query ($k = 20$) by using Kwan [4], with the query itself ranked 1st

									
Rank: 1	2	3	4	5	6	7	8	9	10
D_{query}:0.0	0.61	0.66	0.71	0.74	0.75	0.77	0.78	0.79	0.80
D_{index}:0.0	0.754	0.558	1.526	1.272	2.819	2.211	1.351	1.492	1.623
									
11	12	13	14	15	16	17	18	19	20
0.81	0.82	0.84	0.86	0.866	0.869	0.872	0.875	0.883	0.886
1.630	1.592	1.937	1.548	1.429	1.347	1.491	1.374	1.339	1.573

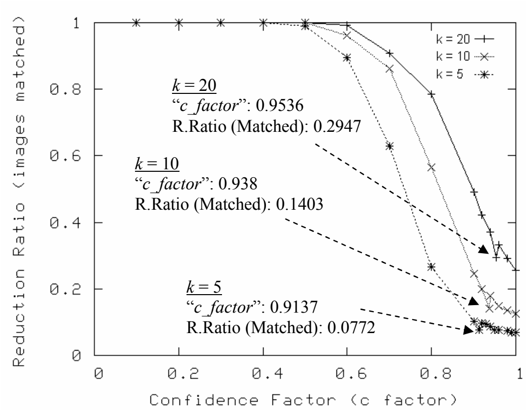


Figure 4: Reduction achieved by filter-and-refine

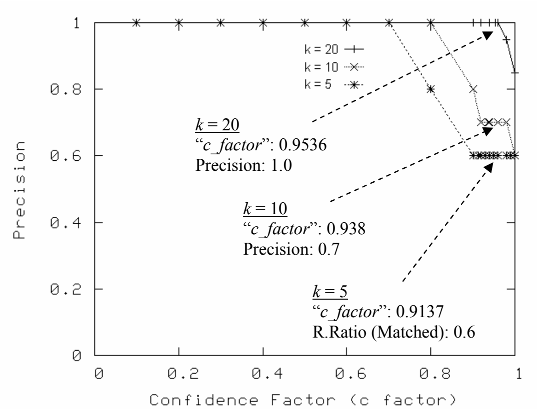


Figure 5: Relation between Precision and Reduction

From Table 4, it is reasonable to conclude that the optimal value *r'* should satisfy this condition, $2.8 < r' \leq 2.9$ for the query and the database that are used in this experiment. At $r = 2.9$, pruning is about 8% while the precision is maintained at 1.0. When we assumed that pruning using a value of $r = 2.9$ has been done before the "filter and refine" steps are performed in the proposed multi-step strategy, the reduction in actual distance calculations is close to 80% for $k = 20$.

Table 4: Precision and reduction ratio by varying *r*

Range (r)	**Precision** (Pr_2)	**Reduction Ratio** (Re_2)
5.0	1.0	1.0
4.0	1.0	0.99
3.0	1.0	0.94
2.9	1.0	0.92
2.8	0.95	0.90
2.6	0.95	0.84
2.4	0.95	0.74
2.2	0.90	0.65
2.0	0.90	0.55
1.8	0.85	0.42
1.6	0.75	0.28
1.4	0.4	0.13

4 Conclusion

In this paper, a novel three-step "prune-filter-refine" strategy for shape similarity search in a Japanese Kamon Image Database is described. First, the "prune" step adopts a spatial index to eliminate improbable matches via an adjustable distance threshold. Second, the "filter" step uses a novel quasi lower-bounding distance derived from a non-metric distance function. Third, the "refine" stage evaluates the remaining candidates by a robust matching method for final similarity ranking. Experimental results confirmed that the proposed strategy achieves larger reduction in actual distance calculations than two-step approaches with close to no false drops in the final retrieval result.

5 References

[1] G.R. Hjaltason and H. Samet, "Properties of embedding methods for similarity search in metric spaces," *IEEE Trans. Patt. Anal. Mach. Intell.* Vol. 25, no. 5, pp. 530-549, 2003.

[2] D. Jacobs, D. Weinshall and Y. Gdalyahu, "Classification with Nonmetric Distances: Image Retrieval and Class Representation," *IEEE Trans. Patt. Anal. Mach. Intell.* Vol. 22, no. 6, pp. 583-600, 2000.

[3] System Product Co. Ltd., "Home Page", http://www.e-spc.co.jp/kamon2/.

[4] P. Kwan, K. Kameyama and K. Toraichi, "On a relaxation-labeling algorithm for real-time contour-based image similarity retrieval," *Image and Vision Computing*, Vol. 21, no. 3, pp. 285-294, 2003.

[5] H. Nagashima, S. Tsubaki and J. Nakajima, "A Classification for Trademark Images Using the Auto-correlation Function Graph Figure," *IEEJ Trans. EIS*, Vol. 123, no. 9, pp. 1547-1554, 2003.

[6] R. Haruki, K. Toraichi and Y. Ohtaki, "A Multi-Stage Algorithm of Extracting Joint Points for Generating Function-Fonts," in *Proc. 2nd Int'l Conf. on Document Analysis and Recognition, Tsukuba, Japan,* pp. 31-34, 1993.

[7] P. Kwan, K. Toraichi, H. Kitagawa and K. Kameyama, "Approximate Query Processing for a Content-Based Image Retrieval Method," In: V. Malik et al.(Eds.): *DEXA 2003*, LNCS 2736, Springer-Verlag, pp. 517-526, 2003.

[8] P. Ciaccia, and M. Patella, "Searching in Metric Spaces with User-Defined and Approximate Distances," *ACM Trans. Database Systems*, Vol. 27, no. 4, pp. 398-437, 2002.