WLSD: A Perceptual Stimulus Model Based Shape Descriptor

Jiatong Li[1], Baojun Zhao[1], Linbo Tang*, Chenwei Deng[1], Lu Han[1], Jinghui Wu[1]
[1] School of Information and Electronics, Beijing Institute of Technology
[e-mail: lijiatong.lee@gmail.com]
*Corresponding author: Linbo Tang
[e-mail: tanglinbo@bit.edu.cn]

Received July 22, 2014; revised October 14, 2014; accepted November 20, 2014; published December 31, 2014

Abstract

Motivated by the Weber's Law, this paper proposes an efficient and robust shape descriptor based on the perceptual stimulus model, called Weber's Law Shape Descriptor (WLSD). It is based on the theory that human perception of a pattern depends not only on the change of stimulus intensity, but also on the original stimulus intensity. Invariant to scale and rotation is the intrinsic properties of WLSD. As a global shape descriptor, WLSD has far lower computation complexity while is as discriminative as state-of-art shape descriptors. Experimental results demonstrate the strong capability of the proposed method in handling shape retrieval.

Keywords: Shape descriptor, shape retrieval, multi-scale representation, feature selection

http://dx.doi.org/10.3837/tiis.2014.12.016

1. Introduction

With the development of the Internet and the rapid update of the electronic equipment, the efficient and effective image retrieval methods are required to satisfy the image retrieval needs of people. Unlike the traditional textual annotation based retrieval methods, content-based image retrieval (CBIR) systems, which represent the image as features like color, texture and shape to implement image matching, are expected to realize the function of searching image in high efficiency [1]. Shape is an important perception for humans to understand images. However, to establish effective shape feature extraction and matching method is still a challenging task. A robust shape descriptor is required to be invariant to translation, rotation and scale as well as insensitive to noise, tolerant to distortion and even occlusion [2]. In addition, the massive database requires the CBIR has quick retrieval response, therefore, low computation complexity is also necessary for a shape descriptor.

Many shape representation and analysis methods have been proposed in the past decades. They can be generally classified into two categories: contour-based methods and region-based methods. Among the two categories, contour-based methods are in the majority.

There are many contour-based methods. One of the classic methods is Curvature Scale Space (CSS) [3]. The CSS uses the zero-crossings of the contour curvature to divide the whole shape contour into convex/concave arcs. CSS smoothens the contour by Gaussian kernels of increasing scales, and finally the whole contour will be convex with pairs of the curvature zero-crossing points evolving together. The whole evolving process of the curvature zero-crossings can be illustrated by the CSS Image, whose similarity is used for shape matching. Another well-known shape descriptor is Shape Context (SC) [4]. At each reference point, SC uses the spatial location of remaining contour points to form the 2D histogram, and thus finds correspondence by the histogram sets of two shapes. In order to tackle the shape articulation, Inner Distance Shape Context (IDSC) [5] extends the SC, and proposes to replace the traditional Euclidean distance by inner distance, which is the shortest path between landmark points within the shape silhouette, and dynamic programming is used to preserve the ordering of the contour points. Unlike SC and IDSC which utilize local histogram matching methods, Contour Points Distribution Histogram (CPDH) [9] uses one histogram to describe the distribution of the whole contour points of a shape under polar coordinate, and the shape similarity is obtained by Earth Mover's Distance (EMD). In [8], another shape descriptor called Contour Flexibility (CF) is proposed, which describes the shape contour points by their deformable potential and also uses dynamic programming to do shape matching. Besides designing discriminative shape descriptors, some new shape matching methods are also proposed. One of them is Hierarchical Procrustes Matching (HPM) [6]. HPM is proposed to conduct the shape matching hierarchically by using the longer segment matching result to predict the correspondence of the shorter and finer segment. In [7], Shape tree (ST) describes the shape by hierarchically dividing the shape contour into short sub-curves, and adopts the elastic matching method for shape comparison. The ST representation has good tolerance to shape distortion, and random shape deformations can be obtained by adding noise to the nodes in a shape tree without perceptually identifying the shape category.

The other major category of shape descriptor is region-based methods. For example, the Zernike moments (ZM) [10] is selected as the MPEG7 standard region-based shape descriptor, but it is relatively time consuming and has limited tolerance to shape distortion. Another moment-based shape descriptor is Moment Invariant (MI) [11]. Although MI

performs not as well as ZM, it only has 7 feature vectors, with low storage and high speed retrieval response, so it is also a widely used shape analysis method. Recently, Support Vector Shape (SVS) [12] is proposed which uses the decision function trained by Support Vector Machine (SVM) to describe the shape. With the Radial Basis Function (RBF), SVS maps points within the shape to high dimension, and has good performance against large noise.

Recent years, there are some new trends in shape analysis. For instance, some work focus on the post-processing after shape matching [13] [14]. In [13], it is proposed to construct the graph model using the similarity (or distance) of pairs of shapes, and adopt the graph transduction to learn the graph structure implicitly. The context-based shape retrieval methods like graph transduction can well handle the situation in which intra-class distances are larger than inter-class ones and is capable to improve the retrieval rate based on available shape measures. In addition, there are researches about the heuristic-based auxiliary shape descriptor aims at describing some specific kinds of shapes. One of the representative work is [15]. In [15], two perceptually motivated strategies are proposed, the first handles shapes with base structure and "strand" structures, the second handles symmetry shapes. The two strategies can be integrated into existing shape matching methods to improve the retrieval or classification performance. In this paper, we focus more on shape representation and matching and do not focus on the above post-processing methods, since they can be used in all available shape measures.

The remainder of paper is organized as follows. In Section 2, we review the previous shape representation work in multi-scale which is in close relation to our approach. Section 3 describes the new shape representation in detail. Section 4 presents the experimental results using the proposed method for shape retrieval in two datasets, including the widely used MPEG-7 core dataset and Tari 1000 dataset respectively. The experimental result in MPEG-7 shows the competitive retrieval performance to the state-of-the-art shape descriptors, and the results in Tari 1000 demonstrate the proposed method is also very discriminative in handling articulated shapes.

2. Related Work

Contour-based methods are more popular than region-based ones and it can further be divided into global description methods and local description methods. Usually, local descriptors are more discriminative than global ones. The massive database requires not only high retrieval rate, but also low storage features and quick retrieval response. However, most local description methods aim to find correspondence of feature sets of contours, consequently sacrificing computation efficiency, which is not suitable for large database retrieval.

Multi-scale methods are an important research branch of shape analysis and have shown high discriminative ability [3][6][7][29][30]. However, most of the existing multi-scale methods still use the scale normalization to get the scale invariance, and circular shift matching to achieve the rotation invariance, all of which add to high computation load. In addition, the general multi-scale shape representation itself is another step that is time-consuming. For example, as discussed in Section 1, CSS achieves the multi-scale representation by smoothing the shape contour using Gaussian kernels. The same representation method is also adopted by [29].

This paper proposes a new contour-based shape descriptor, called Weber's Law Shape Descriptor (WLSD). WLSD is a global shape descriptor with very low computation

complexity while is as discriminative as the state-of-the-art local descriptors. In addition, WLSD is intrinsically extended to multi-scale without extra computation load, and it represents a shape by only 24 dimension vector each scale, which leads to efficient shape matching without scale normalization or circular shift.

WLSD aims to describe the salient variation of a shape contour that stimulate human perception. According to Weber's Law, under the fact that human perception of a pattern depends not only on the stimulus difference but also on the ratio of the relative stimulus change to the original stimulus intensity, we first design Contour Angular Feature (CAF) to model the so-called stimulus intensity, and then establish WLSD according to the Weber's Law Equation. WLSD is then extended to multi-scale to enhance its description capacity. We also adopt the feature selection method to select the effective scales of WLSD. The experiments demonstrate the discriminative power and low computation property of WLSD in shape retrieval.

3. Weber's Law Shape Descriptor for Shape Representation

In this section, we first review Weber's Law and then propose the Contour Angle Feature (CAF). We then adopt the CAF to construct WLSD, and demonstrate the importance of the multi-scale property of WLSD. Finally, we give the WLSD scale selection procedures by SFS.

3.1 Weber's Law

Weber's Law is a psychological rule. It demonstrates that the change in stimulus intensity varies by the original stimulus intensity, and the ratio of the just noticeable difference (JND) of the perception to the original stimulus is a constant [23] [24]. The relationship can be expressed as:

$$\frac{\Delta I}{I} = k \tag{1}$$

where ΔI denotes the JND, I represents the original stimulus intensity, k is the constant corresponding to the certain perception. The fraction $\Delta I / I$ is known as the Weber fraction or Weber proportion. For example, an experiment conducted by Weber in 1840 shows that people can feel the increase or decrease of the additional 1 gram if they hold a 52 grams weight object, but can only feel 2 gram if they hold 104 grams, which demonstrates the JND differs according to the original stimulus intensity.

3.2 WLSD

Weber's Law gives us the motivation to a perceptual stimulus model. The first task is to construct a shape descriptor as the original stimulus intensity. Many shape descriptors have been proposed, such as the distance between the shape centroid and the contour, which is used in [25] to do the trademark retrieval, others, like the contour curvature, signature, Fourier Descriptor (FD), are either difficult to model Weber's Law or are sensitive to scale or rotation variation, so we design a new shape descriptor called Contour Angular Feature (CAF).

3.2.1 The Original Stimulus Feature - Contour Angle Feature

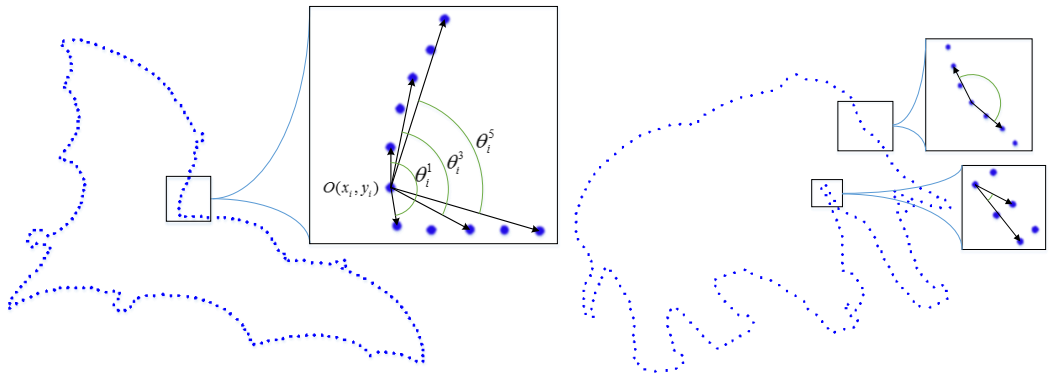
We first introduce the single scale CAF, and then give the general definition of the multi-scale CAF.

Given a closed contour point O with its nearest right and left points A and B, then $\angle AOB$ is called the CAF of point O. Note $\angle AOB$ has two directions, inwards or outwards to the closed contour, and the sum of two angles are 2π. In this paper, we suppose the CAF always be the outwards one. CAF can be extended to multi-scale naturally, let CAF_s signifies the s scale of CAF, then angle $\angle AOB$ is CAF_1, and CAF_2 is the angle formed by contour point O with its second nearest right and left points, etc. Now we give the general definition of CAF_s:

Definition I: Given a closed contour C, and sample it by n clockwise points, whose coordinates are orderly (x_i, y_i), where $i = 1, 2, ..., n$, then the CAF_s of $O(x_i, y_i)$ is defined as:

$$\theta_i^s = \begin{cases} \arccos\left(\dfrac{\overrightarrow{OA_s} \cdot \overrightarrow{OB_s}}{\left|\overrightarrow{OA_s}\right|\left|\overrightarrow{OB_s}\right|}\right) & \overrightarrow{OA_s} \times \overrightarrow{OB_s} > \\ 2\pi - \arccos\left(\dfrac{\overrightarrow{OA_s} \cdot \overrightarrow{OB_s}}{\left|\overrightarrow{OA_s}\right|\left|\overrightarrow{OB_s}\right|}\right) & \overrightarrow{OA_s} \times \overrightarrow{OB_s} < \end{cases} \tag{2}$$

where $s = 1, 2, ..., \lfloor (n-1)/2 \rfloor$ ($\lfloor\ \rfloor$ denotes round toward negative infinity), operator $\times$ represents the vector cross product, $\overrightarrow{OA_s} = (x_{\mathrm{mod}(i+s,n)} - x_i, y_{\mathrm{mod}(i+s,n)} - y_i, 0)$, $\overrightarrow{OB_s} = (x_{\mathrm{mod}(i-s,n)} - x_i, y_{\mathrm{mod}(i-s,n)} - y_i, 0)$.



(a) (b)

Fig. 1. (a) Illustration of the CAF_s
(b) CAF_s varies from nearly 0 (the bottom square) to near π (the upper square)

In **Definition I**, the multi-scale CAF_s can be obtained with s varying from 1 to $\lfloor (n-1)/2 \rfloor$, A_s and B_s are the right and left points relative to O respectively, and both are s points away from O. Consequently, every sample point can have the maximum $\lfloor (n-1)/2 \rfloor$ scale angles. Examples of the CAF_1, CAF_3, and CAF_5 of point $O(x_i, y_i)$ are shown in **Fig. 1** (a). The reason

why extending $\overrightarrow{OA_s}$ and $\overrightarrow{OB_s}$ to three dimension is to highlight the direction property of the vector cross product (for instance, $\overrightarrow{OA_s} \times \overrightarrow{OB_s} > 0$ indicates that $B \to O \to A$ is clockwise ordered), therefore, the definition guarantees angle θ_i^s always directing outwards the interior of a closed contour. Analogously, the CAF_s of the opposite direction can be defined, which is equivalent to Definition I. The defined θ_i^s varies from 0 to 2π, indicating the local contour of the shape changing perceptually from inwards strand structure to outwards one (as illustrated by **Fig. 1** (b)), so the CAF_s is capable to describe all variation of a shape contour. Notice CAF is invariant to scale and rotation.

3.2.2 Weber's Law Shape Descriptor

Hinted by the Weber's Law, we use the CAF difference between the current point and its neighbors to model the difference of the stimulus intensity, and proportion of CAF differences to CAF of the current point to get the saliency variation that stimulate human perception. We then adopt the arctangent function to operate on the proportion [17], which can reduce the negative effect caused by noise. Let θ_i be CAF_1 of the contour point $O(x_i, y_i)$, then the saliency of point $O(x_i, y_i)$ can be described as:

$$f(\theta_i) = \arctan\left(\frac{\theta_{\mathrm{mod}(i+1,n)} - \theta_i}{\theta_i} + \frac{\theta_{\mathrm{mod}(i-1,n)} - \theta_i}{\theta_i}\right) \tag{3}$$

where θ_i denotes the CAF_1 of the current point, $\theta_{\mathrm{mod}(i+1,n)}$ and $\theta_{\mathrm{mod}(i-1,n)}$ represent the right and left neighbor points of the current point respectively. We call this feature the Weber's Law Shape Descriptor (WLSD), and $f(\theta_i)$ the WLSD value of point $O(x_i, y_i)$.

The idea of the WLSD is partially motivated by Weber's Law Descriptor (WLD) in [17]. WLD is an image descriptor regards the gray intensity as the original stimulus intensity, and the difference of the stimulus is the intensity differences of a current pixel and its neighbors. While in WLSD, CAF is viewed as the stimulus intensity, and the difference of the stimulus is the CAF difference of a current contour point and its neighbors. However, the description targets of WLSD and WLD are different, leading to the necessity of miming more information of WLSD: the information content of a shape contour is much smaller than that of a gray image, and the CAF of a contour varies relatively more smoothly than that of gray intensity. That leads to the motivation of extending WLSD to multi-scale.

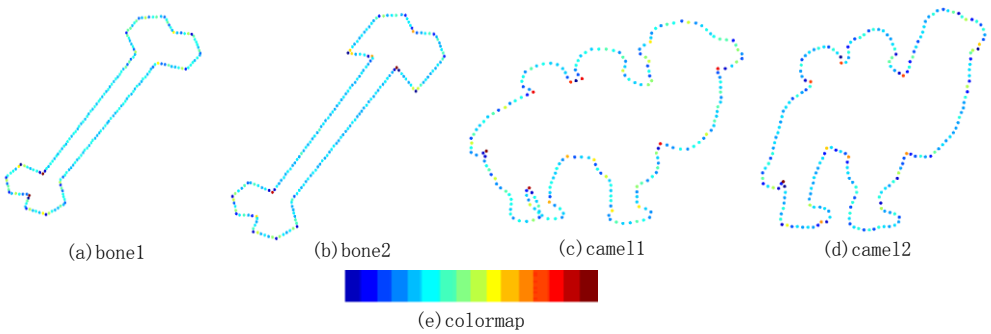


Fig. 2. $WLSD_{1\to1}$ (single scale WLSD) value linearly mapped to the colormap bar

If we only use single neighbor based WLSD shown in Equation (3), the relatively large WLSD value points (salient points) will only occupy a small proportion of the whole contour points. **Fig. 2** maps the WLSD value of contour points linearly to the colormap bar illustrated in **Fig. 2** (e) so that we can observe the description capacity of WLSD visually (the minimum value maps to leftmost deep blue, maximum to the rightmost deep red). **Fig. 2** depicts the single scale WLSD (i.e. the WLSD indicated by Equation (3)) of the contour. Because CAF_1 varies slowly, WLSD value has a sparse distribution, with most of the WLSD value focus on the middle of the colormap bar (nearly 0). In addition, the single neighbor based WLSD value presents an intermittent appearance because of the noise or the continuity variation of the shape contour, which is an unstable and less discriminative description. Our goal is to let WLSD give a relatively continuous and robust description of the contour in accordance to human perception. As a result, we introduce the multi-scale WLSD.

The scale of WLSD has close relation with the scale of CAF, let $WLSD_w$ represents WLSD of scale w, given a contour sampled with n points and its corresponding CAF_s, let $w=1$, then the $WLSD_1$ of point $O(x_i, y_i)$ is:

$$f(\theta_i^s) = \arctan\left(\frac{\theta_{mod(i-s,n)}^s - \theta_i^s}{\theta_i^s} + \frac{\theta_{mod(i+s,n)}^s - \theta_i^s}{\theta_i^s}\right) \tag{4}$$

where θ_i^s signifies the CAF_s defined as Equation (2). Consequently, each $WLSD_w$ has a corresponding scale of CAF_s. In this sense, the scale of WLSD is also determined by the scale of CAF_s implicitly. We use $WLSD_{s\to w}$ represents the $WLSD_w$ extracted from CAF_s, then the notation of the Equation (4) is $WLSD_{s\to 1}$. Combining the scales of s and w, we introduce the general definition of $WLSD_{s\to w}$:

Definition II: Given a closed contour C, and sample it by n clockwise points, whose coordinates are orderly (x_i, y_i), where $i=1,2,...,n$, the $WLSD_{s\to w}$ of point $O(x_i, y_i)$ is defined as:

$$f(\theta_i^s) = \arctan\left(\sum_{k=1}^{w}\frac{\theta_{mod(i-k\times s,n)}^s - \theta_i^s}{\theta_i^s} + \sum_{k=1}^{w}\frac{\theta_{mod(i+k\times s,n)}^s - \theta_i^s}{\theta_i^s}\right) \tag{5}$$

where θ_i^s is the CAF_s of the contour point $O(x_i, y_i)$, and $w=1,...,\lfloor\lfloor(n-1)/2\rfloor/s\rfloor$ ($\lfloor\ \rfloor$ denotes round toward negative infinity).

Therefore, given the CAF_s of a contour, the stimulus intensity difference of the current scale w is the difference between the current point's CAF_s and the CAF_s of its right and left w points (the adjacent distance within each sides' points is s).

Fig. 3 illustrates some examples of $WLSD_{s\to w}$. Red represents the current point, green denotes the right and left w number of points corresponding to different scale $WLSD_{s\to w}$.

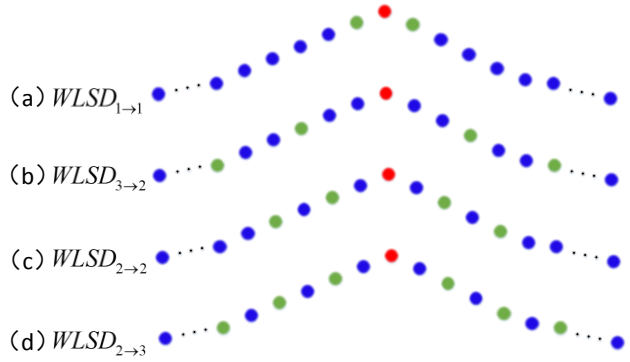


Fig. 3. The illustration of the $WLSD_{s\to w}$

Fig. 4 illustrates that the multi-scale WLSD can describe the contour saliency well and has the desired continuous salient variation description property. (a) ~ (d) depict the $WLSD_{5\to5}$, and demonstrate that multi-scale WLSD has relatively uniform distribution within the WLSD range of value and can describe the contour saliency continuously. **Fig. 4** also indicate WLSD has the similar description of intra-class shapes.

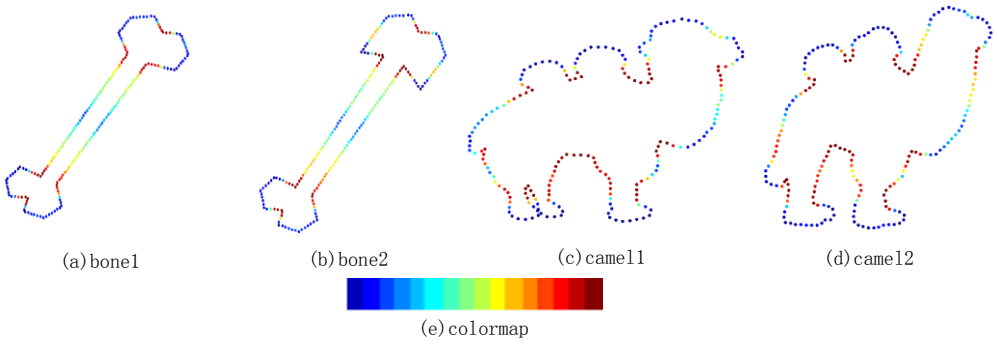


Fig. 4. $WLSD_{5\to5}$ value linearly mapped to the colormap

3.2.3 WLSD Histogram and WLSD scale selection

In order to let the feature be tolerant to the shape distortion to some extent, we uniformly divide the range of $WLSD_{s\to w}$ value, and use one dimensional histogram as the global shape feature. Given two histograms H_1 and H_2, we adopt the commonly used χ^2 distance to compare them:

$$d(H_1, H_2) = \frac{1}{2}\sum_{k=1}^{K}\frac{[h_{1k} - h_{2k}]^2}{h_{1k} + h_{2k}} \tag{5}$$

where K represents the number of histogram bins, h_{1k} and h_{2k} signifies number of points falling into the k th bin of H_1 and H_2 respectively. In the paper we take $K = 24$.

Different scale of WLSD present different discriminative information, for instance, $WLSD_{3\to3}$ can capture local information of the contour, while $WLSD_{16\to5}$ tends to describe the

contour in a global view. As mentioned above, each $WLSD_w$ has a corresponding scale of CAF_s. Given a n points contour, then the whole number of scales of CAF is $\lfloor (n-1)/2 \rfloor$, and each scale of the CAF corresponds to $\lfloor \lfloor (n-1)/2 \rfloor / s \rfloor$ scales of WLSD. Consequently, given a n points contour, we can get the whole number of scales of WLSD, which is denoted by S_{WLSD}:

$$S_{WLSD} = \sum_{s=1}^{\lfloor (n-1)/2 \rfloor} \lfloor \lfloor (n-1)/2 \rfloor / s \rfloor \tag{6}$$

We want to select discriminative scales of WLSD among the whole number of scales shown in Equation (6), and combine the selected scales to be the final effective features. Motivated by the feature selection theory, we view the scale selection as a feature selection problem and use the idea of Sequential Forward Selection (SFS) [36]. First, we define the feature evaluation criterion according to specific application situation, and then select the scale in the candidate set that can improve the defined criterion iteratively until the criterion stops improving. It should be noted that the scales selected by SFS may be not the optimal, but it is relatively more time-saving than many other complicated algorithms, and is very efficient which can be verified by the experiment.

One of the problem above is that, the feature dimension will expand as the iterative times increase, then the shape matching computation will also increase. To overcome the above computation problem, we operate directly on the distance matrix instead of calculating the combined features. Suppose there are M training shapes, then the distance matrix is $\mathbf{D}_{M\times M}$, whose elements are the distance between each pair of shapes. Suppose each contour is uniformly sampled by 200 points, according to Equation (6), the size of the initial candidate set S_{WLSD} is 473, we denote them by:

$$\mathbf{D}_{1\to 1}, \mathbf{D}_{1\to 2}, ..., \mathbf{D}_{1\to 99},$$
$$\mathbf{D}_{2\to 1}, \mathbf{D}_{2\to 2}, ..., \mathbf{D}_{2\to 49},$$
$$...$$
$$\mathbf{D}_{99\to 1}$$

The subscript of $\mathbf{D}$ is in accordance with that of $WLSD_{s\to w}$, it can be simplified denoted by $\{\mathbf{D}_1, \mathbf{D}_2, \ldots, \mathbf{D}_l, \ldots, \mathbf{D}_{S_{WLSD}}\}$. The WLSD scale selection algorithm steps are shown as **Table 1**:

Table 1. WLSD Scale Selection Algorithm

WLSD Scale Selection Algorithm
Initialize: define the evaluation criterion $J(\square)$**;**
calculate the candidate distance matrix set $\{\mathbf{D}_1, \mathbf{D}_2, \ldots, \mathbf{D}_l, \ldots, \mathbf{D}_{S_{WLSD}}\}$**;**
initialize the selected distance matrix set $\mathbf{W}_0 = \varnothing$**, selected scale** $\mathbf{L} = \varnothing$**, and the iterative counter** $k = 0$**;**
while 1
Select the distance matrix $\mathbf{D}_{l_k} = \arg\max(J(\mathbf{W}_k + \mathbf{D}_{l_k}))$ **in the candidate set;**

if $J(\mathbf{W}_k + \mathbf{D}_{l_k}) < J(\mathbf{W}_k)$ **or** $k > S_{WLSD}$ **stop iteration;**

else

add l_k **to the selected scale** $\mathbf{L}$ **;**

update the selected set $\mathbf{W}_{k+1} = \mathbf{W}_k + \mathbf{D}_{l_k}$ **;**

remove $\mathbf{D}_{l_k}$ **from the candidate set;** $k = k+1$**;**

end

end

Obtain the effective scales $\mathbf{L}$ **of** $WLSD_{s\to w}$ **.**

After the above training procedures, the selected scales are recorded in $\mathbf{L}$. In specific application, for example, shape retrieval, we first train the samples to acquire the effective WLSD scales, and then compare shapes according to their combination distances of the selected WLSD scales.

4. Experimental Results and Analysis

In this section, we demonstrate the performance of WLSD in shape retrieval, and compare it with other state-of-art methods. We first do experiment in MPEG7 shape database. Others, like Kimia's dataset, is also very popular in shape matching, but it is not suitable for training since its size is relatively small (Kimia'-99 dataset has 99 shapes with 11 shapes per category). Therefore, we choose the Tari dataset [21]. In the MPEG-7 dataset, we further conduct some specific analysis experiments, including the computation complexity analysis, the stability of the scale selection algorithm and the effectiveness of WLSD in comparison with CAF.

In the following experiments, the retrieval rate is measured by the so-called bull's eye score. Let t be the number of shapes of the same class. Every shape in the database is compared to all other shapes, and the number of shapes from the same class among the $2t$ most similar shapes is reported. The bull's eye retrieval rate is the ratio of the total number of shapes from the same class to the highest possible number (in MPEG7 is 20×1400). We apply the five-fold cross validation to demonstrate the performance of WLSD. The dataset is divided into five folds, one split for validation and the others for training. Note that each validation fold is tested on the whole dataset using the scales obtained from the corresponding four training splits, consequently, the combination of the five validation folds' performance will include retrieval result of all the dataset samples.

Each contour is uniformly sampled by 200 points, according to Equation (6), the number of the initial candidate set S_{WLSD} is 473. The experimental environment is Matlab 7.11.0, and the experimental platform is Intel(R) Core(TM) 2 Duo CPU, 2.53 GHz.

4.1 MPEG-7 dataset

4.1.1 Comparison with existing methods

MPEG-7 CE-Shape-1 Part B dataset is widely used in shape retrieval, it consists of total

1400 shapes, having 70 shape classes, and each class owns 20 pictures. The database is challenging due to the presence of much distortion and examples that are visually dissimilar from other members of their class and highly similar to members of other classes. The first example of each class is shown in **Fig. 5**:

Fig. 5. Examples of MPEG-7 shape dataset

The bull's eye score is compared by the recently popular 13 shape analysis algorithms. From **Table 2**, it can be seen that the bull's eye score of WLSD performs better than all the other methods, including multi-scale methods-[6][29][30]. Moreover, WLSD is much more efficient than all the methods listed on the table. In real application, after the scale selection and the feature extraction of the database having been done offline, the time complexity of shape retrieval using WLSD is only the time costs by feature extraction of the query shape, and the number of selected scales χ^2 distance computation of the 24-dimension vectors. Under the above experimental environment, the pairwise shape matching time of WLSD is only 2.2ms with $n = 200$, where n is the number of contour points.

Considering the methods whose retrieval rate are above 85% listed on the table. Hierarchical Procrustes reports 300ms taken by matching two shapes [6]. The computation of pair-wise similarity between two shapes in Symbolic takes 76.5ms on average, with the contours represented by 100 points [33]. Shape L' Â ne Rouge estimates the density takes on average 2 to 3 minutes per shape, which is also much time consuming than WLSD [7]. With 100 points, IDSC reports 0.31s on a 2.8G PC implemented by optimized Matlab code [5]. The main reasons of time saving of WLSD are two points, one is that WLSD is a global shape descriptor which is much faster than local descriptors that depend on local matching; the other is that WLSD is intrinsically extended to multi-scale without extra extension load, such as filtering the contour used in [18][29].

Table 2. Comparison of bull's eye score on MPEG-7 dataset

Algorithms	***Score(%)***
Hierarchical Procrustes [6]	86.35
Symbolic [33]	85.92
Shape L'Ane Rouge [7]	85.25
IDSC [5]	85.40
Multi-scale Representation [29]	84.93
Polygonal Multi-resolution [30]	84.33
DSW [31]	82.13
Generative Models [19]	80.03
CPDH [9]	76.56
ASD & CCD [32]	76.20
SC [4]	76.51
CSS [18]	75.44
PAF [34]	74.36
WLSD	**86.65**

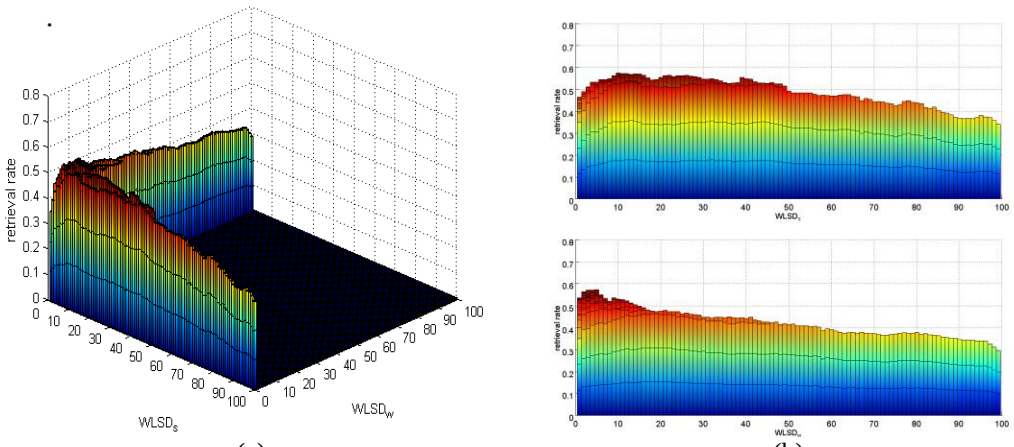
4.1.2 Experiments of the single scale WLSD

We first test the effectiveness of the single scale WLSD by showing its bull's eye retrieval rate of each scale in **Fig. 6** (a). Because of the relation of $WLSD_s$ and $WLSD_w$, the scales mainly spread near the coordinates. The highest score of single scale is 58.30%, with the corresponding scale is $WLSD_{10\to5}$, while lowest score is only 33.16%. The results show that the performance of single scale WLSD is not satisfactory. Consequently, the multi-scale fusion is necessary so that different scales that contain different information about the contour variation can make good combination.

From **Fig. 6** (a), it can be seen that the outstanding scales of WLSD are mainly focused on the low scales relatively, because the score peaks are grouped around the original point. The trends begin to decrease as the scales become larger.

We also want to observe the performance of $WLSD_s$ and $WLSD_w$ respectively. So in **Fig. 6** (b), the front and the left elevation of **Fig.6** (a), the illustration of $WLSD_s$ and $WLSD_w$, are drawn respectively. It illustrates that the outstanding scales of $WLSD_s$ are mainly range from 10 to 40, and the counterpart of $WLSD_w$, which are relatively lower, range from 1 to 20. Moreover, the scores of $WLSD_s$ are obviously higher than those of $WLSD_w$ on average. It is precisely able to explain the characteristic of distribution of the selected scales, which we will show in the following multi-scale experiment.

Although it seems $WLSD_s$ performs better than $WLSD_w$, however, it does not mean that $WLSD_w$ is not important, as it can be seen from **Fig. 6** (a) that only after $WLSD_w$ starts to grow can the $WLSD_{s\to w}$ reaches a jump increasing performance. It proves that $WLSD_s$ and $WLSD_w$ have good combination, and $WLSD_w$ makes necessary supplement to $WLSD_s$.



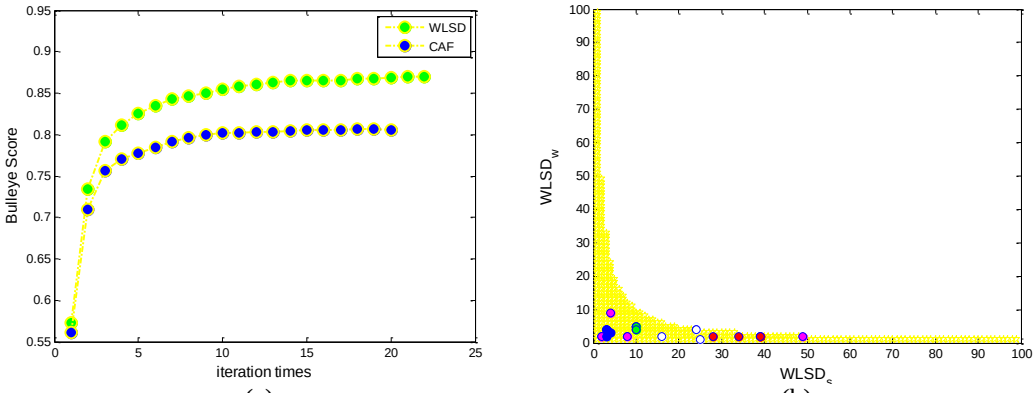
(a) (b)

Fig. 6. (a) The retrieval rate of the single scale WLSD
(b) The front and left elevation of Fig.6 (a)

4.1.3 Experiments of the multi-scale WLSD

In the first experiment in this section, we will demonstrate the effectiveness of the multi-scale integration. First, the detail of the iteration performance is discussed. In **Fig. 7** (a), the curve shows that the retrieval rate increases by iteration using WLSD in MPEG-7. It demonstrates that the power of a single scale WLSD is not very satisfactory but the score increases shapely by 73.41% just after the second iteration. The retrieval rate arrives more

than 80% with less than 4 scales and reaches above 85% with less than 8 scales, which shows the efficiency of the SFS. Finally, after ten iterations, the performance is relatively stable.



(a) (b)

Fig. 7. (a) The improvement trend of the retrieval rate by the increase of the iteration times (b) The first five selected scales of each training set in MPEG-7

Then, we demonstrate the effectiveness of Weber's Law applied to WLSD, the experiment using only the CAF is conducted and the result is also shown in **Fig. 7** (a). The highest score obtained by CAF is 80.59%, which is much lower than WLSD. But the trend of the increasing retrieval rate is similar to that of WLSD. Therefore, WLSD can dramatically improve the discriminative performance compared with only CAF. The main reason is that CAF only has the multi-scale information of a shape contour, whereas WLSD can describe more than that with the saliency variation which implicitly indicates the relation between the contour points. The same increasing trend of the curve shown in **Fig. 7** (a) also indicates that the multi-scale selection motivated by SFS is efficient and stable.

The distribution of the first several selected scales is further observed. **Fig. 7** (b) illustrates the first five selected scales of the five-fold subsets, where the indication color is orderly green, blue, red, white and magenta. It can be seen that the variance of the selected scale becomes larger as the number of iteration increases, and the variance along $WLSD_s$ varies more dramatically than that of $WLSD_w$. So the lower scales of $WLSD_{s\to w}$ are relatively more stable.

In the second experiment, we analysis the WLSD scale distribution. The selected scales of five folds are shown in **Fig. 8** (a) ~ (e) respectively, where the yellow area is all the possible WLSD scales and the green dots represent the selected scales. It is illustrated that the effective scales of $WLSD_w$ are mainly the lower scales between 1~16. Even though the scales of $WLSD_s$ spread relatively scattered, they are more concentrated among 1~20, despite of some sparse high scales. We conclude that the discriminative scales of $WLSD_{s\to w}$ mainly concentrate in lower scales while higher scales make effective complement.

The scales of retrieval rate more than 50% are shown in **Fig. 8** (f) in blue dots. Compared with the selected scales in **Fig. 8** (a) ~ (e), the outstanding individual scales (we call them OIS) are more concentrated. We also find that many selected scales are not in the OIS set although the first selected scales are mainly in the OIS domain. Consequently, less outstanding scales are also very important since they make necessary supplement to the OIS.

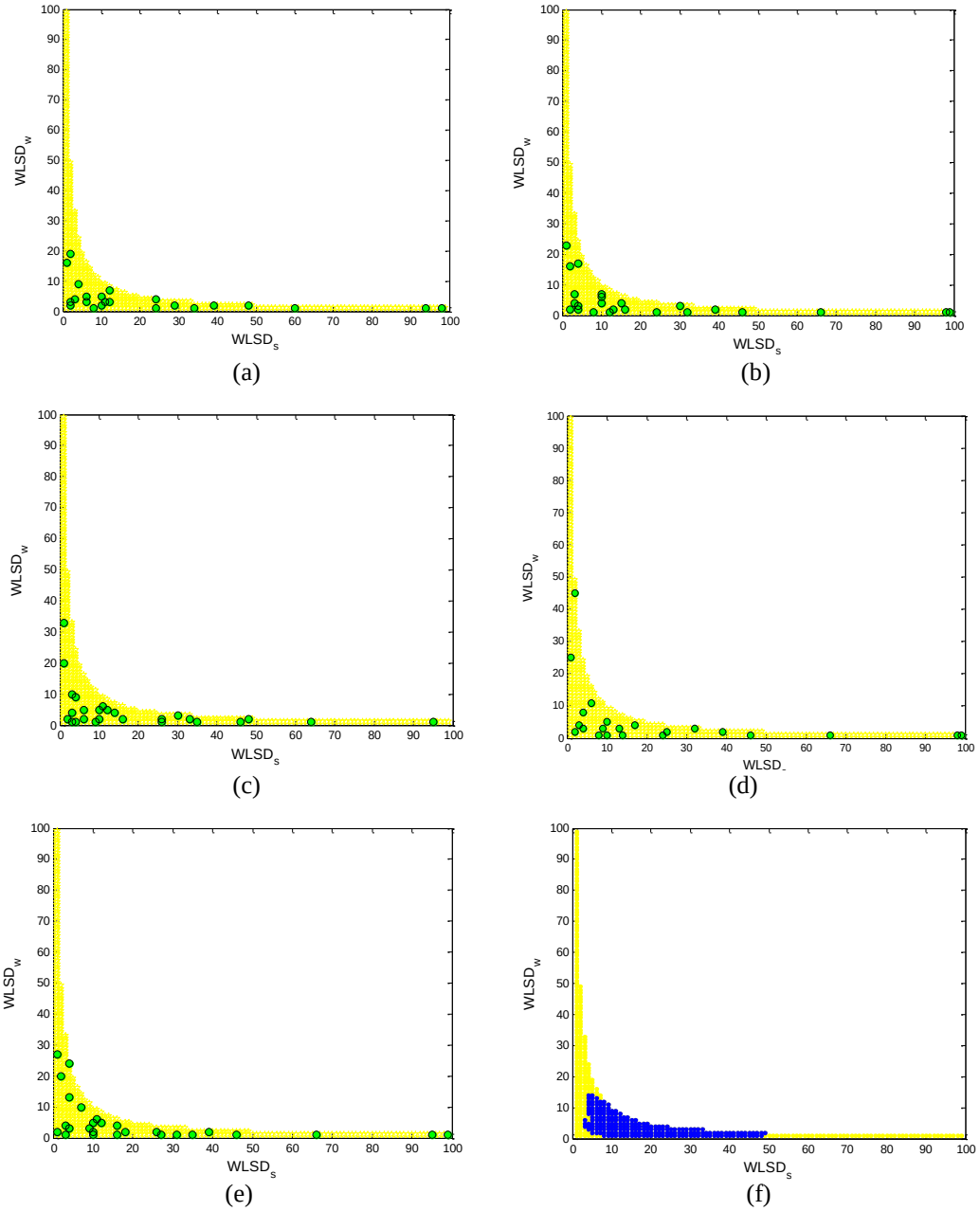
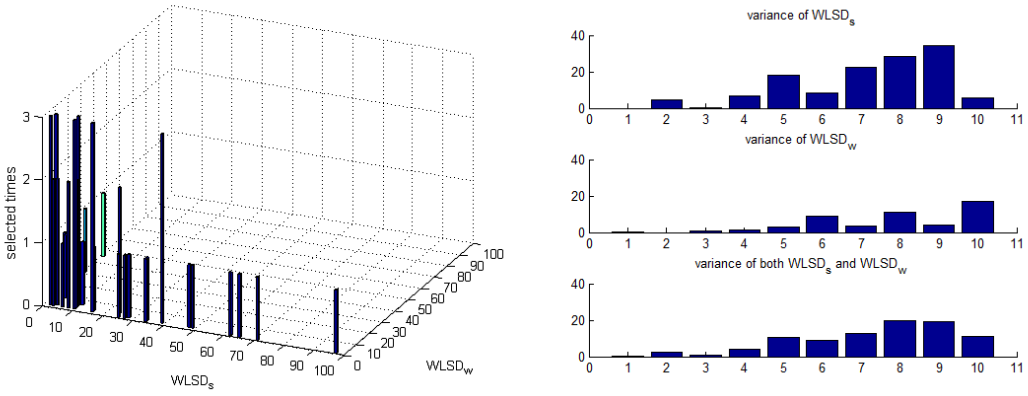


Fig. 8. (a) ~ (e): The distribution of the selected scales of each training set in MPEG-7 (f): The scales of retrieval rate more than 50% in MPEG-7

In the third experiment, we demonstrate the robustness of the proposed scale selection algorithm. In **Fig. 9** (a), the accumulation of the selected scales of the five training sets are illustrated. Compared with **Fig. 7** (b), as more scales are included, the scales of $WLSD_s$ spread more uniformly than $WLSD_w$. It can be explained that once the high $WLSD_s$ scale is selected, the high $WLSD_w$ scale is not necessary to some extent, or in another way, high $WLSD_w$ scales are not as discriminative as the lower ones. We also want to conduct the quantitative analysis of the variation of the selected scales. Consequently, in **Fig. 9** (b), the variance of the first ten

selected scales of the training sets are illustrated. From the first two figures, it can be seen that the variance of the first several selected scales of both $WLSD_s$ and $WLSD_w$ are very stable. Although there are some fluctuations in higher scales, the whole stability is satisfactory. We conclude that the proposed scale selection algorithm is stable in the dominant discriminative scales and have acceptable variation in the complementary scales.

As shown in **Fig. 9** (b), $WLSD_w$ is relatively more stable than $WLSD_s$. Although $WLSD_s$ performs better than $WLSD_w$, $WLSD_w$ has made more contribution in the stability of the whole $WLSD_{s\to w}$. We further draw the conclusion that $WLSD_s$ dominates the discriminability and $WLSD_w$ not only provides necessary supplement but also makes $WLSD_{s\to w}$ more stable. Therefore, it demonstrates again the close relation between $WLSD_s$ and $WLSD_w$.



(a) (b)

Fig. 9. (a) The accumulation of the first ten selected scale of the training sets
(b) The variance of the first ten selected scales

4.2 Tari dataset

To demonstrate the WLSD can also handle articulated shapes, we test it on relatively new Tari dataset, which consists of 1000 binary images from 50 shape categories, each category has 20 images. It is designed to have large intra-class shape deformation and many shapes are articulated (see **Fig. 10**.). It has been extended from the original 180 images [21] to do the shape skeleton research [22].

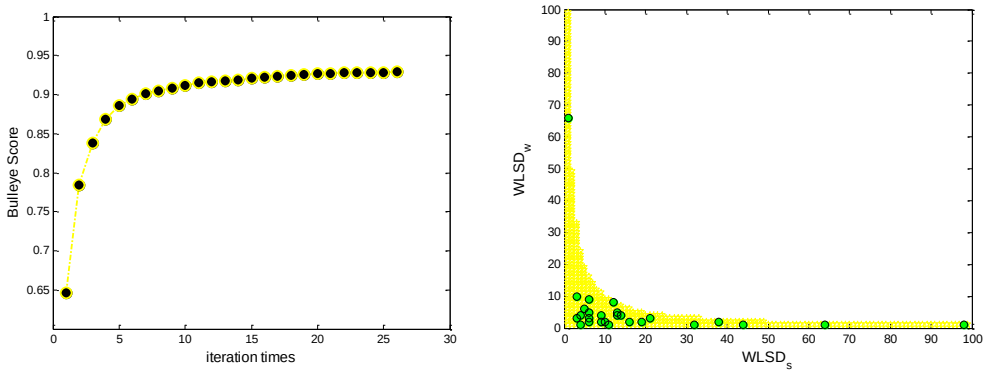
Fig. 10. Examples of Tari dataset

We also adopt the bull's eye score here, since the scores reported in this dataset are not as much as that in MPEG-7, the number of the compared algorithms are relatively small (as shown in **Table 3**).

Table 3. Comparison of bull's eye score on Tari dataset

Algorithm	IDSC [5]	COP [26]	SPM [35]	DSW [31]	**WLSD**
Score(%)	95.33	92.18	91.37	81.60	93.09

From **Table 3**, IDSC scores higher than WLSD since it is designed to handle articulated shapes and is insensitive to articulation. Overall, WLSD is comparable to the listed methods, which prove the proposed method can handle articulated shapes well. Note the compared methods are also much more time-consuming than WLSD since their local property or local matching scheme.



(a) (b)

Fig. 11. (a) The curve shows the retrieval rate increases by iteration in Tari (b) The distribution of the selected scales of WLSD in Tari

The iteration plot and the distribution of scales are illustrated in **Fig. 11**. The curve shows that after the second iteration, the retrieval rate increases significantly and begins approaching to the final rate. The selected scales cluster in lower scales and disperse in the higher scales. Both of the figures are in accordance with the test in MPEG-7 dataset, which demonstrates the discussion and conclusion in MPEG-7 experiment section, and shows the stability of the proposed method.

In **Fig. 12**, it shows that the variance of $WLSD_w$ is much more stable than $WLSD_s$. Compared with **Fig. 9** (b), the average variance of $WLSD_{s\to w}$ here is less variable, which can be explained that the Tari dataset has less distortion than MPEG7. In summary, it has the same variation feature as that of MPEG7. It demonstrates again the robustness of the scale selection method.

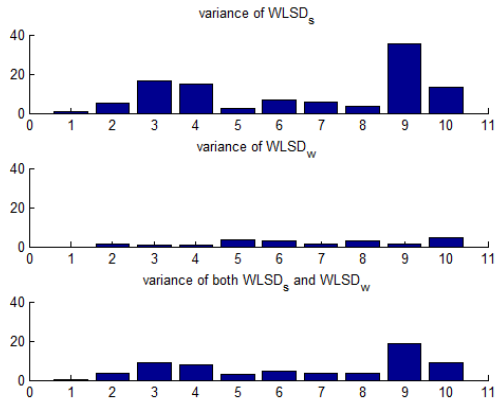


Fig. 12. The variance of the first ten selected scales

As described in the Abstract and Section 3.2.1, both the CAF and WLSD are insensitive to rotation and scale variation, we now demonstrate this property in the last experiment. We

conduct the experiment by rotating the Tari dataset to $30^{\circ}, 60^{\circ}, 90^{\circ}$ clockwise, and scale the dataset to 0.5, 1.5 and 2 as well. The test result is shown in **Table 4**. From the table, it can be seen that the rotated and the scaled scores are nearly the same to the original ones (the rotation 0° and scale 1), which demonstrates CAF and WLSD are robust to rotation and scale variation. We also find that some scores are even higher than the original, which may because of the dataset itself or the de-noising operation in the image transform. The scores of 1.5 and 2 scale transform are lower than that of 0.5 can validate this discussion.

Table 4. Bull's eye score on rotated and scaled Tari dataset

transform	*Rotation*				*Scale*			
	0°	30°	60°	90°	1	0.5	1.5	2
WLSD	93.09	93.27	93.64	93.12	93.09	93.65	93.02	92.83
CAF	87.35	88.86	88.76	87.77	87.35	89.09	87.67	86.35

5. Conclusion

In this paper we propose a new shape descriptor based on Weber's Law, named Weber's Law Shape Descriptor (WLSD). The key idea of WLSD is to capture the salient variation of a shape contour to stimulate human perception. We first design Contour Angular Feature (CAF) to construct the original stimulus intensity and propose WLSD according to Weber's Law Equation, and then develop the WLSD to multi-scale as well as demonstrate the necessity of the multi-scale extension, finally, the feature selection framework is applied to the multi-scale WLSD to extract discriminative scales. The experiments show the effectiveness and the outstanding efficiency of WLSD.

The future work is notable from the theory foundation of WLSD, since Weber's Law is instinctively associated with saliency. Therefore, we plan to enhance the selection scheme to select the salient WLSD that stimulate human views to further improve the discriminative ability.

References

[1] Mussarat, Yasmin, et al, "Content based image retrieval using combined features of shape, color and relevance feedback," *KSII Trans. on Internet and Information Systems*, vol. 7, no. 12, pp. 3149-3165, 2013. Article (CrossRef Link).

[2] Tak, Yoon-Sik, and Eenjun Hwang, "Pruning and matching scheme for rotation invariant leaf image retrieval," *KSII Trans. on Internet and Information Systems*, vol. 2, no. 6, pp. 280-298, 2008. Article (CrossRef Link).

[3] F. Mokhtarian and S. Abbasi, "Shape similarity retrieval under affine transforms," *Pattern Recognition*, vol. 35, no. 1, pp. 31-41, 2002. Article (CrossRef Link).

[4] Belongie Serge, Jitendra Malik, and Jan Puzicha, "Shape matching and object recognition using shape contexts," *IEEE Trans. on Pattern Analysis and Machine Intelligence,* vol. 24, no. 4, pp. 509-522, 2002. Article (CrossRef Link).

[5] Ling Haibin and David W. Jacobs, "Shape classification using the inner-distance," *IEEE Trans. on Pattern Analysis and Machine Intelligence,* vol. 29, no. 2, pp. 286-299, 2007. Article (CrossRef Link).

[6] G. McNeill and S. Vijayakumar, "Hierarchical procrustes matching for shape retrieval," *IEEE Computer Society Conf. on Computer Vision and Pattern Recognition,* vol. 1, pp. 885-894, 2006. Article (CrossRef Link).

[7] P. F. Felzenszwalb and J. D. Schwartz, "Hierarchical matching of deformable shapes," *IEEE*

Conf. on Computer Vision and Pattern Recognition, pp. 1-8, 2007. Article (CrossRef Link).
[8] C. Xu, J. Liu, and X. Tang, "2D shape matching by contour flexibility," *IEEE Trans. on Pattern Analysis and Machine Intelligence,* vol. 31, no. 1, pp. 180-186, 2009. Article (CrossRef Link).
[9] X. Shu and X.J. Wu, "A novel contour descriptor for 2D shape matching and its application to image retrieval," *Image and vision Computing*, vol. 29, no. 4, pp. 286-294, 2011. Article (CrossRef Link).
[10] W. Y. Kim and Y. S. Kim, "A region-based shape descriptor using Zernike moments," *Signal Processing: Image Communication* vol. 16, no. 1, pp. 95-102, 2000. Article (CrossRef Link).
[11] Žunić Joviša, Kaoru Hirota, and P. L. Rosin, "A Hu moment invariant as a shape circularity measure," *Pattern Recognition* vol. 43, no. 1, pp. 47-57, 2010. Article (CrossRef Link).
[12] Van Nguyen Hien, and Fatih Porikli, "Support vector shape: a classifier-based shape representation," *IEEE Trans. on Pattern Analysis and Machine Intelligence,* vol. 35, no. 4, pp. 970-982, 2013. Article (CrossRef Link).
[13] Bai Xiang, et al, "Learning context-sensitive shape similarity by graph transduction," *IEEE Trans. on Pattern Analysis and Machine Intelligence,* vol. 32, no. 5, pp. 861-874, 2010. Article (CrossRef Link).
[14] Yang Xingwei, Suzan Koknar-Tezel and Longin Jan Latecki, "Locally constrained diffusion process on locally densified distance spaces with applications to shape retrieval," *IEEE Conf. on Computer Vision and Pattern Recognition,* pp. 357-364, 2009. Article (CrossRef Link).
[15] Temlyakov, Andrew, et al, "Two perceptually motivated strategies for shape classification," *IEEE Conf. on Computer Vision and Pattern Recognition*, pp. 2289-2296, 2010. Article (CrossRef Link).
[16] Qi Heng, et al, "An effective solution for trademark image retrieval by combining shape description and feature matching," *Pattern Recognition*, vol. 43, no. 6, pp. 2017-2027, 2010. Article (CrossRef Link).
[17] Chen Jie, et al, "WLD: A robust local image descriptor," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 32, no. 9, pp. 1705-1720, 2010. Article (CrossRef Link).
[18] F. Mokhtarian, S. Abbasi, J. Kittler, "Efficient and robust retrieval by shape content through curvature scale space," *Series on Software Engineering and Knowledge Engineering*, vol. 8, pp. 51-58, 1997. Article (CrossRef Link).
[19] Z. Tu, A. L. Yuille, "Shape matching and recognition-using generative models and informative features," *Computer Vision-ECCV*, *Springer Berlin Heidelberg*, pp. 195-209, 2004. Article (CrossRef Link).
[20] Nanni Loris, Sheryl Brahnam and Alessandra Lumini, "Local phase quantization descriptor for improving shape retrieval/classification," *Pattern Recognition Letters*, vol. 33, no. 16, pp. 2254-2260, 2012. Article (CrossRef Link).
[21] Aslan Cagri, et al, "Disconnected skeleton: Shape at its absolute scale," *IEEE Trans. on Pattern Analysis and Machine Intelligence,* vol. 30, no. 12, pp. 2188-2203, 2008. Article (CrossRef Link).
[22] Baseski Emre, Aykut Erdem and Sibel Tari, "Dissimilarity between two skeletal trees in a context," *Pattern Recognition*, vol. 42, no. 3, pp. 370-385, 2009. Article (CrossRef Link).
[23] Gescheider, A. George, "Psychophysics: the fundamentals," *Psychology Press*, 2013. Article (CrossRef Link).
[24] Jain, K. Anil, "Fundamentals of digital image processing," *Prentice-Hall, Inc*., 1989. Article (CrossRef Link).
[25] Qi Heng, et al, "An effective solution for trademark image retrieval by combining shape description and feature matching," *Pattern Recognition*, vol. 43, no. 6, pp. 2017-2027, 2010. Article (CrossRef Link).
[26] Zhou Yu, et al, "Shape matching using points co-occurrence pattern," in *Proc. of 6th International Conf. on Image and Graphics*, pp. 344-349, 2011. Article (CrossRef Link).
[27] H. Ling, X. Yang, and L. J. Latecki, "Balancing deformability and discriminability for shape matching," *Computer Vision–ECCV*, Springer Berlin Heidelberg, pp. 411-424, 2010. Article (CrossRef Link).

[28] Peter Adrian, Anand Rangarajan and Jeffrey Ho, "Shape L'Ane rouge: sliding wavelets for indexing and retrieval," *IEEE Conf. on Computer Vision and Pattern Recognition,* pp. 1-8, 2008. Article (CrossRef Link).
[29] T. Adamek and N. E. O'Connor, "A multiscale representation method for nonrigid shapes with a single closed contour," *IEEE Trans. on Circuits and Systems for Video Technology,* vol. 14, no. 5, pp. 742-753, 2004. Article (CrossRef Link).
[30] Attalla Emad and Pepe Siy, "Robust shape similarity retrieval based on contour segmentation polygonal multiresolution and elastic matching," *Pattern Recognition*, vol. 38, no. 12, pp. 2229-2241, 2005. Article (CrossRef Link).
[31] I. C. Paula, F. N. S Medeiros and F. N. Bezerra. "Shape retrieval by corners and dynamic space warping." in *Proc. of 18th International Conference on Digital Signal Processing*, pp. 1-6, 2013. Article (CrossRef Link).
[32] F. Fotopoulou, G. Economou. "Multivariate angle scale descriptor of shape retrieval," in *Proc. of Signal Process. Appl. Math. Electron. Commun (SPAMEC).* pp. 105-108, 2011. Article (CrossRef Link).
[33] M. R. Daliri, V. Torre, "Robust symbolic representation for shape recognition and retrieval," *Pattern Recognition*, vol. 41, no. 5, pp. 1782-1798, 2008. Article (CrossRef Link).
[34] B. Wang, "Shape retrieval using combined Fourier features," *Optics Communications*, vol. 284, no. 14, pp. 3504-3508, 2011. Article (CrossRef Link).
[35] Z Wang, J. Ouyang, "Shape classes registration and retrieval based on shape parts matching," *Journal of Computational Information Systems*, vol. 9, no. 4, pp. 1493-1499, 2013. Article (CrossRef Link).
[36] L. C. Molina, L. Belanche, À. Nebot. "Feature selection algorithms: a survey and experimental evaluation," *IEEE International Conf. on Data Mining*, pp. 306-313, 2002. Article (CrossRef Link).

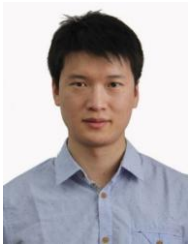
Jiatong Li is a successive Master and Ph.D. candidate in Information and Communication Engineering at Beijing Institute of Technology (BIT) under the supervision of Professor Baojun Zhao. He is now visiting University of Technology, Sydney (UTS) as a dual-doctoral degree student under the supervision of Richard Xu. His research interests include pattern recognition and machine learning.

Baojun Zhao received the Ph.D. degree in electromagnetic measurement technology and instrument from Harbin Institute of Technology (HIT) in 1996. From 1996 to 1998, he worked as a Postdoctoral Research Fellow at Beijing Institute of Technology (BIT). Baojun Zhao is currently a professor and the director of National Professional Laboratory of Signal Acquisition and Processing. He has been the project leader of more than 30 national projects. His main research interests include video coding, pattern recognition, infrared/laser signal processing, and parallel signal processing. He has authored or co-authored over 100 publications and received 5 national ministerial-level scientific & technological progress awards in these fields.

Linbo Tang received the Ph.D. degree in signal and information processing from Beijing Institute of Technology in 2005. Currently, he is a tutor of master students at the School of Information and Electronics, Beijing Institute of Technology (BIT). His current research interests include video coding, target detection, pattern recognition and object tracking. His speciality is real-time image processing.

Chenwei Deng received the Ph.D. degree in signal and information processing from Beijing Institute of Technology in 2009. Currently, he is an Associate Professor at the School of Information and Electronics, Beijing Institute of Technology. Prior to this, he was a Post-doctoral Research Fellow with the School of Computer Engineering, Nanyang Technological University. He has authored or co-authored over 50 technical papers in refereed international journals and conferences, and co-edited one book. His current research interests include video coding, quality assessment, perceptual modelling, feature representation and object recognition.

Lu Han received her BS degree in Communication Engineering at North Eastern University (NEU) in 2008. She is now pursuing for her MS degree in Information and Communication Engineering at Beijing Institute of Technology (BIT). Her research interests include real-time image processing and pattern recognition.



Jinghui Wu received his BS degree in electronic and information from School of Information and Computer engineering, Northeast Forestry University (NFU), Harbin, China, in June 2007. He received his MS degree in signal and information processing from School of Electrical and Electronic Engineering, Harbin Science and Technology University (HSTU), Harbin, China, in June 2010. He is currently pursuing for his Ph.D. degree in Signal and Information Processing at the School of Information and electronic, Beijing Institute Technology (BIT). His research interests include image enhancement, target tracking, and target recognition.