Width Distributions for Shape Description

Xiaozheng Zhang[†‡], Yongsheng Gao[‡†]

[†] Biosecurity Group, Queensland Research Lab, National ICT Australia
{paul.zhang, yongsheng.gao}@nicta.com.au

[‡] Computer Vision and Image Processing Lab, Griffith University, Australia
{x.zhang, yongsheng.gao}@griffith.edu.au

Abstract—**Measuring the similarity between articulated shapes is a fundamental yet challenging problem. This paper proposes a novel shape descriptor based on Width Distributions (WD), which is robust to articulations. We show that the width distributions are articulation insensitive yet descriptive to distinguish different shapes with varied part structures. With measurements on distributions only, the proposed method does not require any alignment between two objects and thus is more robust than correspondence-based measurements. First, the medial axes of the objects are extracted and fitted to B-Splines to remove outliners. The width of the object shape perpendicular to the medial axis can be calculated for each position on the medial axis. The histograms of those widths are compared using chi-square method for similarity. Experiments on standard 2D shape database show that the proposed method performed better than standard shape distribution algorithms and similarly to other articulation-robust shape descriptors, such as inner-distance. The speed of the proposed method is much faster than the more sophisticated inner-distance, as the proposed method is much simpler in nature and requires limited image processing and pattern classification. These results suggested it could be an efficient and effective method to describe and to match articulated shapes.**

Keywords-**shape, width, medial axis, distrubutions, articulations.**

I. Introduction

Shape descriptions are one of the fundamental tasks for image processing and computer vision. After the object extraction in 2D images or data acquisition of 3D models in 3D space, human being tends to describe the extracted/obtained object based on either their shapes and/or surfaces properties (sometimes, surface properties can be roughly expressed as colors or textures dependent on the primary features of the surfaces). For shape descriptions, human beings tend to using geometric primitives to approximate the overall or partial outlooks of the object. In computer vision, it is extremely useful to describe shapes for different purposes, such as pattern classification (based on shape differences), shape retrievals, shape recognition, shape reconstruction or recovery [13], shape encoding, tracking [5], image compression, etc. For instance, face shapes play an importance role in pose-invariant face recognition, as they are the carriers of facial features, especially for 3D face recognition [14-16].

The task of a shape descriptor is to maximise the differences of different (between-class) shapes, while tolerating the differences between similar (within-class) shapes. The difficulty, however, often lies to the balances between the two objectives. The within-class differences are often caused by issues other than shape differences, which could be capturing variations, image noises, and/or imperfections of preprocessing algorithms. For instance, capturing of a book from different viewing angles may result in different aspect ratios. Also, shape detection or edge extraction should include or exclude wrong pixels. With these issues, the shape descriptors' performances could be compromised. In many applications, the objects are flexible, which make the shape descriptions even more difficult. These objects may be human bodies with arms and legs articulating, animal bodies such as insects' bodies, tools, hands (potentially useful for hand signal understanding), etc. A rigid shape descriptor will very likely describe the shape variations caused by articulations as the intrinsic shape differences, hence gives incorrect results. To tolerate articulations in shape descriptions, local attributes such as part signatures are encouraged over global attributes. Hence, histograms and/or distributions are very useful for this task of articulation robust shape description.

In this paper, we propose a novel shape descriptor, namely the Width Distribution (WD) descriptor, which is robust to articulations. This shape descriptor records the local width information on medial axis as a distribution for similarity measures. In this way, the global structural information is minimised and the local attributes can be utilised as the main shape descriptive features. To measure similarity between two WDs, a standard histogram comparison using Chi-square algorithm is used. The experimental results show some encouraging performances with standard articulated tools image database.

The remainder of the paper is organised as follows. The next section provides a summary of related works. An overview of the proposed approach is presented in Section 3 as well as technical details. In Section 4, experiments on standard shape database and results are reported. Finally, Section 5 contains a summary of this paper with conclusions and future perspectives.

II. RELATED WORKS

The problem of determining the similarity of two shapes is an important and challenging problem in the field of computer vision. A few review papers covered the entire topic quite adequately, such as [7, 11] etc. In this section, we first categorise the shape descriptors with articulation tolerance into 3 categories. Then, two representative approaches closely related to our approach will be discussed in detail.

To tolerate articulation in shape descriptions, three main approaches were/are being adopted. The first approach is so called example-based approaches, which exhaustively save all the possible articulation cases for each shape and use these examples for matching. A few reported works within this category include [2, 12] etc. These approaches typically need large storage space which is not only a hardware challenge, but also a challenge for algorithmic speed. Due to exhaustive sampling, shape variability cannot be too high as this would exponentially increase the storage and computation requirements.

The second category is to align the two shapes with different or the same articulations, which requires correspondence between two shapes. These correspondences were usually established by comparing the local regions [6] or topological properties of simplified shape (e.g., skeleton-based correspondence matching [10]). The correspondence establishment typically involves a rigid matching (even the region is local), which violates the conditions of articulations for joint points and these joint points are usually more important than other "ordinary" points. Also these approaches are slow because they usually involve a local searching and fine-tuning.

The third category is distribution-based approaches, which are very popular in shape matching without considering articulation tolerance, such as [9], as well as other fields such as local binary patterns for texture and face matching [8], etc. Such approaches use statistics on some local attributes to describe the intrinsic features of the targets, without the requirement of properly aligning them or sampling the articulation instances. Therefore, they are computationally much more economical and are more robust to misalignment or mis-sampling. However, most of the distribution-based approaches are sensitive to shape variations. For instance, local binary patterns are sensitive to head rotations and expressions, which can be interpret as head shape variations. To find a distribution-based shape descriptor which is robust to articulation-type shape variations requires a choice of non-global attributes to be recorded for distributions. The proposed method belongs to this category and we choose the width distributions along medial axis as the features for distributions, for the purpose of articulation toleration.

One of the representative methods in the shape matching robust to articulation based on alignment is inner-distance [6]. Ling et al. proposed a novel descriptor, called inner-distance, which is similar to geodesic distance in 2D space and travels within the object shape to connect the shortest distance between to contour points. As shown in Fig. 1, the inner distances of two corresponding points in two shapes varied by articulations remain almost unchanged (dash line), while the

direct distances (or normal distances) can be very different (red line). Based on this articulation robustness, a shape context descriptor using inner distance was proposed and comparing to shape context using Euclidean distance, it performed better in articulated shapes.
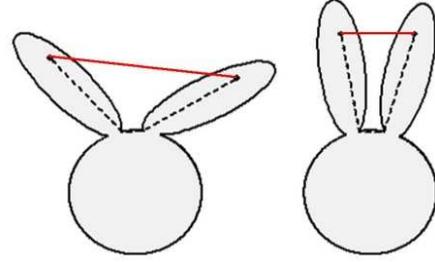

Figure 1. Inner-distance (dash) vs. Euclidean distance (red) on articulated shapes [6].

A representative work in the third category is shape distributions, which used distance distributions for shape matching. In [9], Osada et al. proposed to build distance distributions of different shapes and to use these distributions for shape similarity measurements. The most successful algorithm from all of the proposed shape distributions is the so-called D2 method, which record distances between any pair of two boundary points and build the distance histogram as shown in Figure 2. Using a mean histogram measurement, the two shapes can be compared and a similarity score was calculated. However, this work was for rigid objects only and didn't provide the mechanism to tolerate articulation or other shape variations.
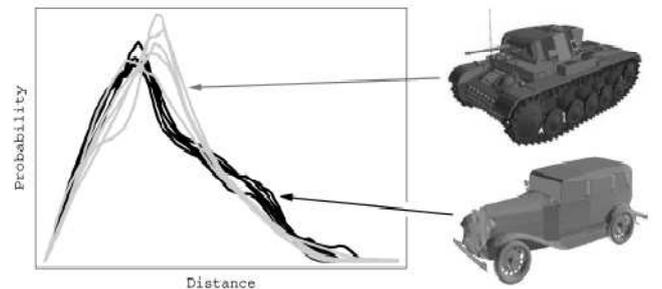

Figure 2. D2 distributions of different 3D shapes [9].

Comparing to the inner distance, our proposed approach doesn't require alignment and provides articulation tolerance in a more efficient way. It is similar to D2 in the way of capturing distributions as features for shape matching, but provides articulation tolerance.

III. WIDTH DISTRIBUTIONS

It is desirable for a good articulation robust shape descriptor to provide descriptive features free from articulations. In this paper, the proposed width distribution approach captures the widths of object shape along medial axis which are insensitive to shape variations caused by articulations. Under the assumption that most parts of articulated shapes remain unchanged and articulation only

affects local areas, the width distributions won't be affected much under articulations.

A. Width distribution calculation

In Figure 3, a shape is first binarised and then skeletonised to extract medial axis. The medial axis was fitted to a B-Spline using a standard fitting algorithm as used in [18], in order to remove the effect of boundary noises of the shape. The width distributions are recorded in the histogram as shown in Figure 4. Such raw width distributions are easily affected by noise and rounding errors. So the raw width distributions are smoothed using the Gaussian filter expressed as

$$d'(w) = g * d(w), \tag{1}$$

where

$$g = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(d-\mu_d)^2}{2\sigma^2}} \tag{2}$$

and quantised using

$$\hat{d}'(u) = \sum_{w=u*s}^{(u+1)*s-1} d'(w), \tag{3}$$

where s is the quantisation range. The quantised width distribution is shown in Figure 5.
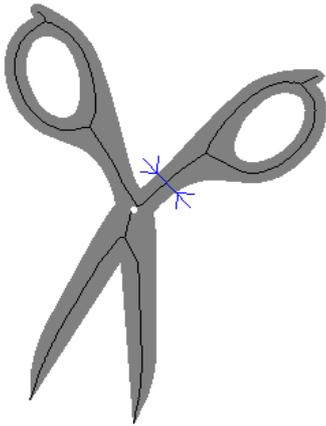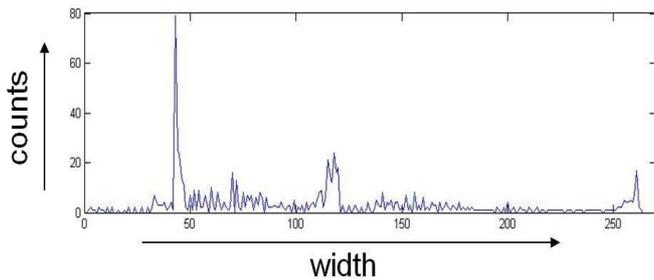

Figure 3. Width extraction of a shape.

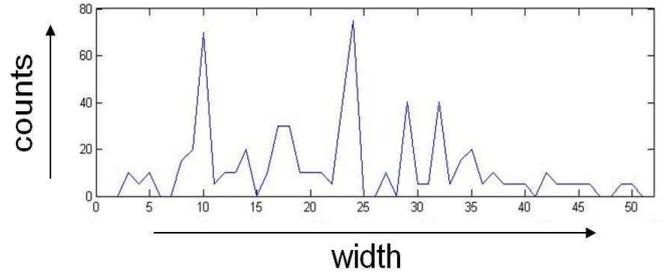Figure 4. Raw width distributions of Figure 3.

Figure 5. Width distribution quantisation.

B. WD's articulation tolerance

To illustrate the property of articulation tolerance of the proposed WD approach, two comparison experiments were conducted where one is overlaying the WDs of the same shape under articulations (Figure 6) and the other is overlaying the WDs of different shapes (Figure 7). The shapes involved were shown on top of each WD chart. Comparing to Figure 7, WDs of shape under different articulations shown in Figure 6 demonstrated much higher correlations and thus can be used as shape descriptor. On the other hand, the WDs of different shapes are not correlated to each other as shown in Figure 6. It is observed from this illustration that the proposed WD approach is insensitive to articulations while maintaining the features of different shapes. Therefore, it is a good choice for the task of shape descriptions with articulation toleration.
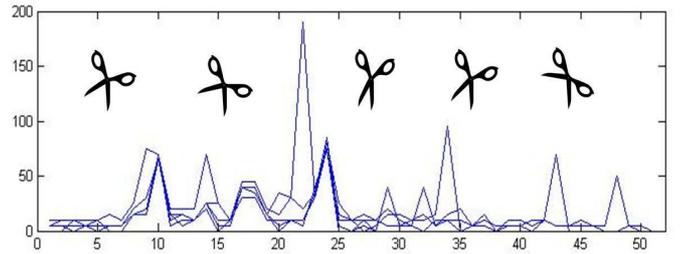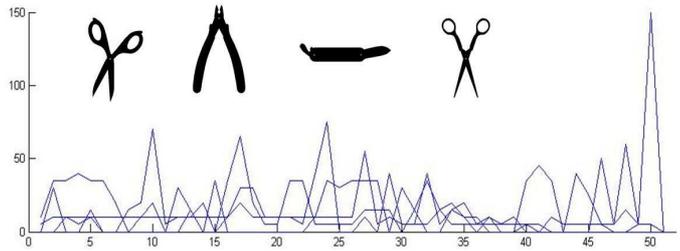

Figure 6. WDs of shape under articulations.

Figure 7. WDs of different shapes.

C. WD's simularity measure

The WDs can be seen as histograms for similarity measurements and there are many histogram measurements in the literature. The comparison of different histogram measurements (matching) is certainly out of the scope of this

paper and interested readers can refer to the comparison reports such as [4] for that purpose. In this paper, we chose to use chi-square as the histogram comparison algorithm. Please note, the proposed WD approach is not limited to chi-square method as used in [1] and we believe that the choice of different histogram similarity measures will only perturb the performance of the proposed WD approach, but not affect the general idea and shape description capabilities.

Denote the two WDs as $\hat{d}_i$ and $\hat{d}_j$, the chi-square measure is calculated [3] as

$$\chi^2(\hat{d}_i, \hat{d}_j) = \frac{1}{2} \sum_{w=1}^{K} \frac{[\hat{d}_i(w) - \hat{d}_j(w)]^2}{\hat{d}_i(w) + \hat{d}_j(w)}, \qquad (4)$$

where K is the total number of width quantisation slots. The lower the chi-square value, the more similar the two WDs are.

IV. EXPERIMENTS

To evaluate the proposed WD approach, a standard tool dataset was used as used in [6] and shape retrieval experiment was conducted. The tool dataset contains totally 35 images of 7 different shapes (all are tools) with 5 different articulations for each shape. The thumbnails of all of the images are shown in Figure 8. The shape retrieval used the strategy of one-against-all, which selected one image as the query and compared it against all of the rest images. The retrieved images are ranked by the similarity measures which are the negative chi-square values. The imagery results of a few examples are shown in Figure 9 and the overall shape retrieval performance of the proposed WD approach was summarised in Table I, with comparison with D2 [9] and inner-distance [6].
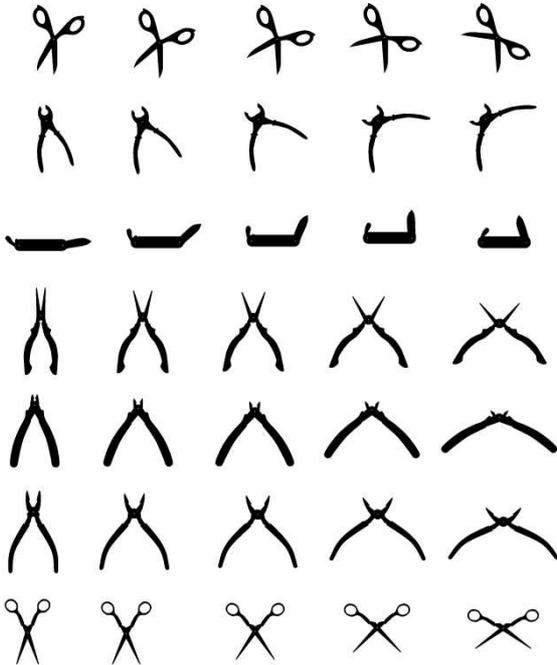

Figure 8. Tools image database.

Figure 9 shows some retrieval results of the proposed WD approach on the tools database. 5 of the correct retrievals are shown on the left columns and 2 of the incorrect retrievals are shown on the right columns. The proposed WD approach generally performs well on shapes with articulations, especially the shapes with variable width along medial axes such as scissors with shape knife tips. However, it has difficulties sometimes in handling the shapes with unchanging widths such as the knife or spanner with uniform bodies. Of course, one can combine the alignment of medial axis and the width distributions to maintain the sequential information, but in the price of slowing the algorithm down and the risk of misalignments.
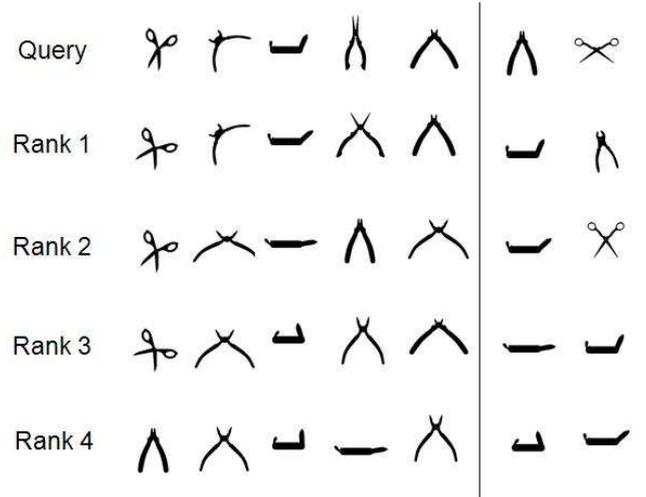

Figure 9. Imagery retrieval results of Tools database.

TABLE I. RETRIEVAL PERFORMANCE OF D2, INNER-DISTANCE, AND THE PROPOSED WD APPROACHES

Method	Rank 1	Rank 2	Rank 3	Rank 4
D2	10/35	14/35	17/35	18/35
Inner-distance	32/35	34/35	34/35	35/35
WD	**29/35**	**32/35**	**33/35**	**34/35**

Comparing to the D2 descriptor, which is not designed to tolerate articulations, but to measure general rigid shapes, the proposed WD approach performed much better on the tools database which contains primarily articulation variable shapes. To the D2 descriptor, different objects and different articulations all mean different shapes. The proposed WD approach, however, could differentiate with an 82% rank 1 accuracy the object differences and articulation differences. With rank 4 measurement, the proposed WD approach only missed one of the 35 retrievals.

The inner-distance approach, which is primarily designed to tolerate articulations, performed slightly better than the proposed WD approach. In rank 4 measure, it achieved 100% accuracy. This is probably because the additional alignment-based shape context provides additional descriptions of the shapes. However, with real-world situations where every instance of the object could be slightly different, the alignment could be much more difficult to be accurate. Moreover, comparing the running times as shown in Figure 10, the

proposed WD approach and D2 approach are much faster than the inner-distance approach, due to the lack of the time-consuming alignment steps.

TABLE II. THE SPEED COMPARISONS OF D2, INNER-DISTANCE AND WD APPROACHES

Method	D2	Inner-distance	**WD**
Run time	7.43 sec	3 min 35 sec	**8.92 sec**

Table II shows the running times of the D2, inner-distance, and the proposed WD approaches using Matlab codes on a Dell Latitude E6500 laptop with Intel Duo Core 2.66GHz CPU and 3.48GB RAM at 1.57GHz memory speed. The run times are for tools images of 635 x 476 pixels and for the entire 35 x 34 comparisons. From Table II, it is obvious that the proposed WD approach is much faster than that of the inner-distance. With the reported performances, one cannot conclude that the proposed WD approach is superior than inner-distance as the retrieval performance is slightly lower. However, it is obvious that it shows the promises of high speed with a small compromise of descriptive power, which is a fairly good trading-off for shape descriptions.

V. CONCLUSIONS

In this paper, a novel shape descriptor has been proposed, the Width Distributions (WD), which is robust to shape articulations. The proposed WD approach extracts the width of the shape along medial axis and uses the histograms of the widths for shape similarity measures. Using a popular chi-square histogram measure, the proposed WD approach was able to differentiate different shapes with the existences of shape articulations, which is a desirable feature for matching flexible objects such as human bodies, tools, animals, etc. In experiment, a standard tools database was used to evaluate the capability of the proposed WD approach, where the database contains a set of shapes with articulations. Comparing to two state-of-the-art approaches, the WD approach showed its ability to tolerate articulations well and performed efficiently. Therefore it is a promising approach for matching articulated shapes.

The future work will be extending this approach to 3D shape matching, as we believe it would be able to differentiate 3D shapes with the existences of articulations. The only difference is to use section areas to replace width in the distributions. It is also interesting to combine the width distributions with the shape context as in the inner-distance approach [6] or colour information as in [17] to achieve higher performance with sacrifice of the speed.

REFERENCES

[1] T. Ahonen, A. Hadid, and M. Pietikäinen, "Face description with local binary patterns: application to face recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 12, pp. 2037-2041, 2006.

[2] P. F. Felzenszwalb and D. P. Huttenlocher, "Pictorial Structures for Object Recognition," *Int. J. Comput. Vis.*, vol. 61, no. 1, pp. 55-79, 2005.

[3] P. E. Greenwood and M. S. Nikulin, *A guide to chi-squared testing*. Wiley, New York, 1996.

[4] W. Jia, H. Zhang, X. He, and Q. Wu, "A Comparison on Histogram Based Image Matching Methods," in *IEEE International Conference on Video and Signal Based Surveillance*, 2006.

[5] H. Li, C. Shen, and Q. Shi, "Real-time visual tracking with compressed sensing," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 11)*, 2011.

[6] H. Ling and D. W. Jacobs, "Shape Classification Using the Inner-Distance," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 2, pp. 286-299, 2007.

[7] S. Loncaric, "A survey of shape analysis techniques," *Pattern Rec.*, vol. 31, pp. 983-1001, 1998.

[8] T. Ojala, M. Pietikainen, and T. Maenpaa, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 7, pp. 971-987, 2002.

[9] R. Osada, T. Funkhouser, B. Chazelle, and D. Dobkin, "Shape distributions," *ACM Transactions on Graphics*, vol. 21, no. 4, pp. 807-832 2002.

[10] H. Sundar, D. Silver, N. Gagvani, and S. Dickinson, "Skeleton based shape matching and retrieval," in *Shape Modeling International*, pp. 130 - 139, 2003.

[11] R. C. Veltkamp and M. Hagedoorn, "State-of-the-Art in Shape Matching," *Principles of Visual Information Retrieval*, pp. 89-119, 2001.

[12] I. Weiss and M. Ray, "Recognizing Articulated Objects Using a Region-Based Invariant Transform," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 10, pp. 1660- 1665, 2005.

[13] X. Zhang, Y. Gao, and M. K. H. Leung, "Multilevel quadratic variation minimization for 3D face modeling and virtual view synthesis," in *Proc. Int. Multimedia Model. Conf.*, pp. 132-138, 2005.

[14] X. Zhang, Y. Gao, and B.-L. Zhang, "Recognising rotated faces from two orthogonal views in mugshot databases," in *Proc. ICPR*, vol. 1, pp. 195-198, 2006.

[15] X. Zhang, Y. Gao, and M. K. H. Leung, "Recognizing rotated faces from frontal and side views: an approach towards effective use of mugshot databases," *IEEE Trans. Info. Foren. Sec.*, vol. 3, no. 4, pp. 684-697, 2008.

[16] X. Zhang and Y. Gao, "Face recognition across pose: a review," *Pattern Rec.*, vol. 42, no. 11, pp. 2876-2896, 2009.

[17] X. Zhang, Y. Gao, and T. Caelli, "Colour Adjustment and Specular Removal for Non-Uniform Shape from Shading," in *Australian Conference on Digital Image Computing: Techniques and Applications*, pp. 563-568, 2010.

[18] X. Zhang, Y. Gao, and T. Caelli, "Primitive-based 3D Structure Inference from a Single 2D Image for Insect Modeling, Towards an Electronic Field Guide for Insect Identification," in *11th. Int. Conf. Control, Automation, Robotics and Vision*, pp. 866-871, 2010