Strategies for Shape Matching using Skeletons

Wooi-Boon Goh

School of Computer Engineering,
Nanyang Technological University,
Nanyang Avenue
Singapore 639798

Email: aswbgoh@ntu.edu.sg
Phone: 65-67904611
Fax: 65-67926559

Abstract

Skeletons are often used as a framework for part-based shape description and matching. This paper describes some useful strategies that can be employed to improve the performance of such shape matching algorithms. Firstly, it is important that ligature-sensitive information be incorporated into the part decomposition and shape matching processes. Secondly, part decomposition should be treated as a dynamic process in which the selection of the final decomposition of a shape is deferred until the shape matching stage. Thirdly, both local and global measures must be employed when computing shape dissimilarity. Finally, skeletal segments must be weighted by appropriate visual saliency measures during the part matching process. These saliency measures include curvature and ligature based measures. Experimental results show that the incorporation of these strategies significantly improves shape database retrieval accuracy.

Keywords: Shape Matching, Skeletons, Medial Representation, Multiresolution
1. Introduction

Medial or skeleton-based approaches to shape analysis have been widely used and proven to be very effective in representing and classifying 2D shapes [6], [9], [24], [26], [27], [29]. This is because skeletons are very amenable to part-based (local) description of shapes and it has been acknowledged that part-based representation is necessary in order to handle common shape matching problems such as occlusion, limb growth and articulation, part addition and deletion [16], [26]. Additionally, Kimia [17] provides a comprehensive discussion of the possible role of medial geometry in the human visual system. Neurophysiological experiments by Lee et al. [19], [20] observed peak neural activity in the V1 cells at localities of medial axis and psychophysical experiments such as those by Kovacs [18] suggest that medial points may be used in encoding shape information.

Unfortunately, a common criticism of medial-based representation of shapes is that skeletonization algorithms are sensitive to small boundary deformation and ligatures [1]. Ligatures are problematic because they can appear and disappear with non-visually salient changes in the shape (see Fig. 4a). If not properly handled, such unstable skeletal features can alter skeletal topology and skew the matching cost between visually similar shapes [32]. Some like van Eede et al. [34] addressed this problem with the use of canonical skeletons that eliminates unstable medial branches while retaining only salient part structure of shape. It is not necessary that ligatures have to be removed before shape matching, but it is important that ligature-sensitive information be strategically incorporated into the shape matching algorithm as is discussed in this paper.

The use of a part-based approach to handle problems such as occlusion requires a shape to be decomposed into meaningful parts. This approach gives rise to another problem, that is, how should a shape be decomposed into parts? Assuming that a shape has now been decomposed into it component parts, another question arises. How should the parts of one shape be matched to another? Is local part-to-part matching sufficient or should shapes also be compared based on a more global geometric descriptor? This paper addresses these questions with some useful strategies that can be incorporated into part-based shape matching algorithms that employ skeletons. In order to test their efficacy, the strategies presented are implemented within the multiresolution gradient vector field (MGVF) skeleton framework of [11], [13], which is briefly introduced in section 2. An
important feature of the MGVF framework that is relevant to this work is the availability of a ligature-sensitive measure in the form of a vector field directional disparity map. Sections 3, 4, 5 and 6 discuss these strategies and their respective implementation. Section 7 presents experimental results, which supports the performance improvement resulting from the adoption of these strategies. Fig. 1 provides the reader with an overview of the entire part-based shape matching process. It also serves as a guide which describes how the various aspects of the proposed strategies presented in this paper come together at various stages of the shape analysis process.
Fig. 1. An overview of the entire part-based shape matching process. It starts with skeleton extraction, followed by dynamic part decomposition and then the matching of the various part-based descriptors using appropriate visual saliency weighting.

2. The MGVF Skeleton

The MGVF skeleton is extracted from a gradient vector field as shown in Fig. 2. This vector field is generated within a multiresolution framework as described in [11], [13]. The property of an extracted skeleton varies with the skeletonization algorithm employed and some properties may be fundamentally at odds with each other [5]. For example, skeletons that have high noise immunity are usually smooth and not plagued by undesirable perturbation-induced runners. However, such skeletons violate the mediality requirement needed for complete reconstruction. What constitute a desirable property is ultimately dependent on the nature of the application. The MGVF skeleton was designed specifically for use in part-based description and matching of silhouette shapes. Properties such as noise immunity, scale and rotation invariance exhibited by the MGVF skeleton are considered desirable for such purposes. Another characteristic the MGVF skeleton shares with medial representation such as the ‘cores’ of Pizer et al. [23] is the disconnected nature of its skeleton (see Fig. 2b). Unlike most skeletonization algorithms [19], [30], the MGVF skeletonization algorithm is specifically designed such that it does not enforce connectivity and this leads to two types of skeletons, namely structural and textural skeletons [13]. It has been argued that this dual representation is desirable for a noise robust medial representation and fast shape matching [11], [13].

![MGVF Skeleton](image)

Fig. 2. (a) The multiresolution vector gradient field from which, (b) the MGVF skeleton is extracted. The structural skeleton represents general shape topology while disconnected textural skeletons capture prominent boundary artifacts (e.g. key’s serrated teeth pattern).

2.1 Vector Field Directional Disparity Map
Given the gradient vector field \( \mathbf{v}(x) \) as shown in Fig. 2a, a skeleton can be located by detecting locations in the vector field where the local gradient vectors exhibit maximum directional disparity. This is similar in concept to the technique proposed by Siddiqi et al. [30], where the measure of average outward flux over a very small neighborhood was used to detect medial points. In this work, the idea of Ben-Arie and Wang in [4] is extended to give an alternative local directional disparity operator \( \Delta \) given by

\[
\Delta(\mathbf{v}(x), \sigma_D) = \frac{g_{\sigma_D}(x) \| \mathbf{v}(x) \| - g_{\sigma_D}(x) \| \mathbf{v}(x) \| \cdot \mathbf{v}(x)}{g_{\sigma_D}(x) \| \mathbf{v}(x) \|}
\]  

(1)

where \( x = (x, y) \) is a 2D point on the image and the normalized disparity measure \( \Delta(\mathbf{v}(x), \sigma_D) \in [0, 1] \) is computed within a weighted locality defined by a Gaussian convolution kernel given by

\[
g_{\sigma_D}(x) = \frac{1}{\sqrt{2\pi}\sigma_D} \exp\left(-\frac{x^2 + y^2}{2\sigma_D^2}\right)
\]

(2)

The local directional disparity operator \( \Delta \) gives a value close to 1 at the center of a converging gradient vector field, such as one obtained at the center of a circle and a value close to 0 in a locality where the gradient vector field is unidirectional, as shown in Fig. 3.

Fig. 3. The output of the local directional disparity operator \( \Delta \) in different vector fields. (a) In a converging vector field, \( \Delta \to 1 \) and (b) in a unidirectional vector field, \( \Delta \to 0 \).

Details of the MGVF skeletonization algorithm are beyond the scope of this paper and can be found in [13]. However, what must be highlighted is the vector field directional disparity map \( M(x,y) \) given by equation (14) in [13]. This disparity map is obtained with
the aid of the directional disparity operator $\Delta$ given in (1) along with some additional processing. The directional disparity measure $M_R(x,y)$, along the skeleton correlates well with the measure of ligature stability (see Fig. 4b). The ligature-sensitive information available in the MGVF is employed in the various strategies discussed. Other measures for ligatures such as those characterized by high negative curvature on the generating shape boundary segment as proposed by Torsello and Hancock [32] may also be employed.

![Diagram](image)

Fig. 4. (a) Unstable ligatures may appear with minor changes to the shape, resulting in a corresponding change to the skeletal topology. (b) Such ligatures have low values in the directional disparity map $M_R(x,y)$.

Three basic strategies for improving shape matching algorithms are discussed. The first in section 3 is related to the process of part decomposition. The second in section 4 discusses appropriate visual saliency measures associated with skeleton-based description of shapes. The third in section 5 argues for the use of both local and global measures when comparing parts of a shape. Underlying the first two strategies is the earlier acknowledgement that ligature-sensitive information must be brought to bear when parts of a shape are being decomposed and when their dissimilarity cost is being computed during part matching. The manner in which this is being implemented can be observed by the use of the directional disparity map $M_R(x,y)$ in sections 3 and 4. Not all the strategies suggested can be readily incorporated into existing skeleton-based matching algorithms, but their general principle should be noted so that existing algorithms can be made more robust by employing as many of these principles as possible within their respective implementation framework.

3. Dynamic Part Decomposition
Traditionally, shape skeletons are first decomposed into parts based on some set of rules [2], [25]. Dissimilarity between shapes is computed by finding the optimal combination of matches between these parts. However, part decomposition is a dynamic process that depends on the two shapes being compared. The intuitive decomposition of shape A in Fig. 5a is different when it is forced to match shapes B or C. Furthermore, shapes can deform in a continuous manner, making it difficult to decide how they should be decomposed during different deformation transitions until one knows the shape it is compared to (see Fig. 5b).

![Image](image1.png)

Fig. 5. (a) Part decomposition is dependent on the shape being matched. (b) Inconsistent skeleton segmentation (part decomposition) based on the local maxima-minima of the vector field directional disparity measure $M_D(x,y)$ along the skeleton with only minor changes to the width of the protruding limb. This may cause the dissimilarity measure between shapes B and C to be much higher than between shapes B and A.

Dynamic part decomposition is implemented by first segmenting the shape’s skeleton into a set of $NP$ basic skeletal components called primary skeletons, given by $\{SP_k\}_{k=1}^{NP}$. Various segmentation criteria [14] such as skeletal junction, maxima-minima of curvature and directional disparity along the skeleton (see Fig. 5b) were used to decompose the shape into all possible component parts (see examples in Fig. 6). Segmentation at prominent curvature and disparity profile extremas help identify probable locations of limb joints, where articulation can occur.
Fig. 6. Segmentation of some MGVF skeletons into a set of primary skeletons \( \{SP_k\} \).

A merging process is then employed to re-group the numerous primary skeletons into a set of \( NM \) visually continuous and coherent merged segments, \( \{SM_m\}_{m=1}^{NM} \). The merging process follows after the notion of visual conductance that Katz and Pizer [15] used to derive natural parts-hierarchy. Two criteria decide which primary skeleton pair at a branch point should merge. Like in [15], the first is the orientation continuity between two primary skeletons \( SP_i \) and \( SP_j \) connected at branch point \( b \) given by

\[
CO_b(SP_i,SP_j) = \frac{|\pi - |\theta_i(b) - \theta_j(b)| |}{\pi} \in [0,1] \tag{3}
\]

where \( | | \) gives the absolute value and \( \theta_i(b) \in [0,2\pi) \) is the local orientation estimated about point \( b \) using the subset of consecutive skeletal points on \( SP_i \) as shown in Fig. 7.

Fig. 7. Orientation continuity between a pair of skeletal segments at the branch point \( b \) is computed using their respective local orientation estimates at \( b \). Values of \( N_{CO}=4 \) was used to provide a robust local orientation estimate for \( 129 \times 129 \) pixel-sized images.

The second measure of visual continuity uses ligature-sensitive information available in the directional disparity map \( M_b(x,y) \) as shown in Fig. 8. Notice that medial points with high degree of substances [15] correspond to the skeletal points with large disparity values and unstable ligatures [1] maps well to points with high degree of
connection and low disparity values. In order to compute the disparity continuity between two adjoining skeletal segments, their respective local mean direction disparity $ED_k(b)$ at branch point $b$ is first computed as shown in Fig. 8. The disparity continuity between two primary skeletal segments $SP_i$ and $SP_j$ connected at branch point $b$ is therefore given by

$$CD_b(SP_i, SP_j) = \frac{|ED_i(b) - ED_j(b)|}{ED_i(b) + ED_j(b)} \in [0, 1] \quad (4)$$

Medial points with high degree of substance
Medial points with high degree of connection

Fig. 8. Measuring the local mean directional disparity value of the skeletal segment at the branch point gives information similar to the substance and connection measures proposed in [15]. Values of $N_{ED} = 10$ were used to compute the mean disparity $ED_k(b)$ of primary skeleton $SP_k$ at branch point $b$.

Since both the orientation continuity and disparity continuity measures give small values when the visual continuity between the two adjoining skeletal segments is favorable, a combined measure for the visual conductance between two primary skeletons $SP_i$ and $SP_j$ connected at branch point $b$ can be formulated using

$$VC_b(SP_i, SP_j) = \gamma CO_b(SP_i, SP_j) + (1 - \gamma)CD_b(SP_i, SP_j) \in [0, 1] \quad (5)$$

where the visual continuity emphasis parameter $\gamma \in [0, 1]$ is used to determine whether orientation or disparity continuity should be the more dominant merging criterion. The skeleton pair with the lowest value of $VC_b$ is merged. The disparity continuity criterion is important as it prevents unstable ligatures from interfering with the decomposition process (see Fig. 9a). Notice in Fig. 9b, the both horses’ backbone skeletons were consistently re-grouped despite initial separation due to ligatures.
Fig. 9. (a) Skeleton merging at Horse 1’s head highlights how using disparity continuity (CD) avoids topological interference by ligatures despite the strong orientation continuity (CO) at the junction. (b) Appropriate disparity continuity emphasis of $\gamma=0.1$ (obtained empirically) resulted in similar merged skeletons for Horses 1 and 2.

A merged segment $SM_m$ can be decomposed into an exhaustive set of meaningful combination of connected primary skeletons as shown in Fig. 10a (this set includes individual primary skeletons). Each of this valid combination of primary skeletons is termed a secondary skeleton, $SS_j$. The permutations of these secondary segments that reconstruct the merged skeleton describe all permitted part decomposition options available for that merged skeleton (see Fig. 10b). Notice that appropriate merging of segmented skeletons serves to reduce the large combinatorial possibility of part decomposition. Furthermore, in the presences of unpredictable topological changing ligature formation, the disparity-based visual continuity criterion encourages the selection of part decomposition permutations that are visually consistent over different shapes that share similar gross visual forms (see Fig. 9b). The dynamic decomposition strategy presented here is similar in principle to graph topology altering algorithms such as [9], [26], [29], [31] that use operations such as insert, merge, delete, contract to accommodate deformation arising from occlusion and limb growth. Giblin and Kimia [10], for example, derived six different types of instabilities that allow appropriate edit distances to be computed during graph matching operations such as those employed in [26], [31]. However, the strategic use of ligature-sensitive information described here helps reduce the number of possible decomposition option and therefore speed up the shape matching process. All visually meaningful part decomposition options are explicitly extracted at this stage for each shape. The preferred decomposition option is not decided until during
shape matching when similar information regarding the shape being matched is also available. Discussion on this selection process is deferred till section 6.

Fig. 10. (a) All possible secondary skeletons extracted from a merged segment $SM_m$ consisting of an ordered set of $N_m=3$ primary skeletons. (b) The $2^{N_m-1}$ possible combinations to reconstitute the merged segment gives all valid part decomposition options for $SM_m$.

4 Incorporating Visual Saliency

Not all parts of a shape are perceived with equal significance. It is therefore important that each extracted part should be associated with some visual saliency measure that can be used to weigh its contribution during shape comparison. There are several characteristics in the skeletal-based part that contributes to its visual saliency. In this work, three saliency measures are explored, namely directional disparity-based saliency, length-based saliency and curvature-based saliency.

4.1 Directional Disparity-based Saliency

Since parts are formed by skeletal segments, the visual saliency of a part must in some measure be related to the formation stability of such skeletons. Notice in Fig. 11b, that the formation stability of segment $Y$ is much weaker compared to that of segment $X$ as the shape undergoes minor transformation. In other words, the formation of segment $Y$ vanishes with a slight change to the boundary contour. Such an unstable skeletal segment should be associated with a low visual saliency measure. As seen in Fig. 11a, segments with weaker formation stability have correspondingly lower values in the directional disparity map $M_R(x,y)$ and this permit a normalized disparity-based saliency measure to be computed for each secondary skeleton $SS_j$ and it is given by
\[ GD(SS_j) = \frac{DS(SS_j)}{\sum_{k=1}^{NP} DS(SP_k)} \in [0,1] \]  

where \( DS(SS_j) \) gives the total directional disparity \( MR(x,y) \) summed over all points in secondary skeleton \( SS_j \) and the denominator gives the total disparity of the shape’s skeleton, which is essentially defined by the complete set of \( NP \) primary skeletons, \( SP_k \).

Fig. 11. (a) Unstable ligatures are associated with the low disparity values \( MR(x,y) \) along skeletal segment (e.g. segment Y). (b) The application of a disparity saliency will reduce the mismatch cost associated with missing segment Y when matching shapes A and C.

### 4.2 Length-based Saliency

Another obvious measure of a skeletal part’s saliency is given by the normalised length of its associated skeletal segment. As suggested by Siddiqi and Kimia [28], a more extended limb is usually considered visually more salient than a shorter limb. The normalized length-based saliency measure of a secondary skeleton \( SS_j \) is given by

\[ GL(SS_j) = \frac{LS(SS_j)}{\sum_{k=1}^{NP} LS(SP_k)} \in [0,1] \]  

where \( LS(SS_j) \) gives the total number of points in each secondary segment \( SS_j \) and the length of each secondary skeleton is normalized by the total skeletal points in the shape’s skeleton. The combined saliency for a secondary skeletal segment \( SS_j \) is given by

\[ G(S_j) = \frac{1}{2} \left[ GD(SS_j) + GL(SS_j) \right] \in [0,1] \]
4.3 Curvature-based Saliency

An attribute of a skeletal-based part that is often used during part-based matching is the curvature of the skeleton [26]. However, Fig. 12a and 12b suggest that the mismatch in skeletal curvature does not necessarily mean two shapes are dissimilar, implying that the curvature-based visual saliency is not constant along a skeletal-based part. This saliency variation is essentially due to the limb’s width along the skeleton (i.e. radius of the maximal disk centered about a skeletal point). It is important that curvature-based dissimilarity measures between two skeletal segment be computed using curvature saliency weighting along its length (see Fig. 15).

Fig. 12. Contribution of skeletal curvature to shape dissimilarity is dependent on the limb width profile along the skeleton. (a) Similar shapes with dissimilar skeletal curvature. (b) Dissimilar shapes that are differentiate by their differing skeletal curvature.

5 Combining Local and Global Shape Measures

Part descriptors employing only local features cannot capture the coherent structure of a shape. Based on local measures alone, dissimilar shapes A and B in Fig. 13a are considered good matches as all parts in one have a local representation in the other. On the other hand, we can easily tell apart the two categories of spectacles and dumb-bells in Fig. 13b even though they share similar local parts. The observed categorization can only be realized by incorporating measures that take into account the global inter-part relationships within the shape. But purely global measures have drawbacks too as they are sensitive to limb articulation, as seen in Fig. 13c. A robust shape dissimilarity measure must therefore consist of a weighted combination of both local and global part-based matching costs. In this work, the former is termed part distance and the later is termed global distance, and their derivation within the MGVF skeleton framework is described.
Fig. 13. (a) Two dissimilar shapes that share identical local part features. (b) Global geometry helps classify shapes sharing similar parts into different categories. (c) Part articulation is problematic for shape descriptors using only global geometry.

5.1 Part Distance

The part distance between two secondary skeletons $SS_i$ and $SS_j$ gives the weighted combination of several error measures that are computed by comparing several part descriptors of $SS_i$ with the corresponding ones from $SS_j$. The local part descriptors used comprise of part-based orientation histograms [12], [14], limb width profile and skeletal curvature along the skeleton. Each of their related error measures is briefly described.

Fig. 14. Relevant attributes of a skeletal segment $SS_j$ consisting of an ordered set of skeletal points $p_j(1)$ to $p_j(N_j)$. These include the source node, $p_j(1)$, the sink node, $p_j(N_j)$, the local skeletal orientation $\theta_j(i)$ and the maximal disk about skeletal point $p_j(i)$ that encloses a set of points given by the set $\{C_j^i\}$. The end of a skeletal segment that is assigned the source node must be the end closest to the shape’s centroid, provide it is a segmentation point (e.g. branch point).
Orientation histogram error – Shapes have been described by vector fields [4] and orientation histograms [22], [33] derived from vector fields. With the availability of a vector field in the MGVF skeleton framework, histograms of the gradient vectors’ orientation within the interior regions of a part are used as local part descriptors. Three orientation histograms are associated with each skeletal part, namely the segment, source node and sink node histograms. As highlighted in Fig. 14, a skeletal segment SS\(_j\) consisting of an ordered set of \(N\) skeletal points \(p_j(\mathbf{i})\), where \(p_j(1)\) is termed the source node and \(p_j(N)\) the sink node. The segment histogram \(H_{sg}^j\) describes the general shape of the part spanned by the skeletal segment \(SS_j\). A curvature-invariant description of the part is formed by cumulating rotation-normalized orientation histograms along the skeletal segment. The orientation histogram at skeletal point \(p_j(\mathbf{i})\) is extracted by finding the set \(\{C_j^\mathbf{i}\}\) of all discrete image pixels within the circle circumscribed by the maximal disk radius \(r_j(\mathbf{i})\) that is centred about \(p_j(\mathbf{i})\). The orientation histogram \(H^j_\mathbf{i}\) consisting of \(n\) bins representing the value range \([0, 2\pi]\) is formed by cumulating the quantized orientation associated with the gradient vector of each pixel within \(\{C_j^\mathbf{i}\}\). The biases resulting from skeletal bending and shape rotation are removed from the orientation histogram by adding the local skeletal orientation \(\theta_j(\mathbf{i})\) to all orientation values before cumulation. Since broader sections of a limb are visually more salient than narrower sections, the normalized orientation histogram is weighted by the radius \(r_j(\mathbf{i})\) of the maximal disk at \(p_j(\mathbf{i})\). The resulting limb width-weighted \(n\)-bin orientation histogram at \(p_j(\mathbf{i})\) is given by

\[
h^j_\mathbf{i}(k) = r_j(\mathbf{i}) \frac{H^j_\mathbf{i}(k)}{\sum_{b=1}^n H^j_b(b)}
\]

The normalized segment histogram \(H_{sg}^j\) for skeletal segment \(SS_j\) with \(N\) points is then obtained by summing all the limb width-weighted orientation histograms computed along each point on the segment and is given by

\[
H_{sg}^j(k) = \frac{\sum_{i=1}^{N_j} h^j_i(k)}{\sum_{i=1}^{N_j} \sum_{b=1}^n h^j_b(b)} \quad \text{where} \quad \sum_{k=1}^n H_{sg}^j(k) = 1
\]

The choice of bin size \(n\) is discussed in Appendix B. The source and sink node histograms can be obtained by computing the local orientation histograms at the respective segment end points but this approach fails for tapering limb extremities that have very small end point maximal disks. Instead end point histograms are obtained by cumulating orientation
histograms over several additional skeletal points from each of the respective source and sink nodes. In short, the $n$-bin orientation histogram at the source node is given by

$$H_{sr_j}(k) = \frac{\sum_{i=1}^{u} h^i_j(k)}{\sum_{i=1}^{u} \sum_{b=1}^{n} h^i_j(b)}$$

where $u = \arg \max_{1 \leq u \leq N_j} \left[ \sum_{i=1}^{u} DS_j(i) \leq \rho \sum_{i=1}^{N_j} DS_j(i) \right]$ (11)

where $h^i_j$ is given in (9) and $DS_j(i)$ give the directional disparity value at point $p_j(i)$ on segment $S_j$. The number of skeletal points involved in computing the source node histogram is given by the largest index value of $u$ that will ensure a fraction $\rho$ of the total directional disparity along $S_j$ is obtained. The sink node histogram is given by

$$H_{sn_j}(k) = \frac{\sum_{i=v}^{N_j} h^i_j(k)}{\sum_{i=v}^{N_j} \sum_{b=1}^{n} h^i_j(b)}$$

where $v = \arg \min_{1 \leq v \leq N_j} \left[ \sum_{i=v}^{N_j} DS_j(i) \leq \rho \sum_{i=1}^{N_j} DS_j(i) \right]$ (12)

The disparity fraction parameter $\rho \in [0,1]$ for computing a end point histograms was set at $\rho = 0.1$ or $\rho = 0.3$ depending on whether the skeletal end point is a connection or terminus point (see Fig. 14). Human visual system-based motivation for this devising this scheme is detailed in [14]. The orientation histogram error between two secondary skeletal segments $SS_i$ and $SS_j$ is the weighted matching cost of their segment, source node and sink node orientation histograms and is given by

$$d_H(SS_i, SS_j) = \frac{\eta_1 d_\chi(Hsg_i, Hsg_j) + \eta_2 d_\chi(Hsr_i, Hsr_j) + \eta_3 d_\chi(Hsn_i, Hsn_j)}{\eta_1 + \eta_2 + \eta_3}$$

(13)

where $d_\chi(\cdot)$ is the $\chi^2$-dissimilarity measure used to compared histograms [3]. The weight $\eta_i$ regulates the contribution of each histogram type to the overall histogram error and the division by the sum of these weights ensures the histogram error $d_H(SS_i, SS_j) \in [0,1]$. Values of $\eta_1 = 1.5$ and $\eta_2 = \eta_3 = 1.0$ were used in the experiments presented.

Limb width error - A useful part descriptor is the function describing the variation in limb width of the part as one traverses the secondary skeletal segment $SS_j$ from source node to sink node. The limb width information is readily available in the maximal disk radii given by $r_j(i)$, as seen in Fig. 14. A set of $r$ polynomial coefficients $\{R_j\}$ can be used to efficiently represent the normalized limb width profile along $SS_j$ and is given by

$$\{R_j^k\}_{k=1}^r = \text{polyfit} \left[ \frac{i}{N_j}, \frac{r_j(i)}{N_j}, \frac{r_j(i)}{N_{\text{norm}}^j} \right]^{-1}$$

(14)
where the operator \texttt{polyfit}[x,y]^n computes the coefficients of a polynomial function \( f(x) \) of degree \( n \) that fits the data \( f(x(i)) \approx y(i) \), in the least-squares sense. The segment length \( N_j \) normalizes the \( x \) values to unit length and the value \( N_{\text{norm}} \), given by the square root of the shape’s total area, normalizes the \( y \) values. The limb width error between two secondary skeletal segments \( SS_i \) and \( SS_j \) can then be efficiently computed using just their \( r \) polynomial coefficients and is given by

\[
d_R(SSI, SS_j) = \left| \sum_{k=1}^{r} \left( \frac{1}{k} \{R_i^k - R_j^k\} \right) \right|
\]

(15)

where \( | | \) gives the absolute value of the summation.

**Skeletal curvature error** - Another useful descriptor of a part is its skeletal bending characteristics. This can be obtained by unwrapping the ordered local skeletal orientation \( \theta_j(i) \) along the secondary skeletal segment \( SS_j \) and is given by

\[
\{\Theta_j(i)\}_{i=1}^{N_j} = \text{unwrap}[\{\theta_j(i)\}_{i=1}^{N_j}]
\]

(16)

where the \texttt{unwrap}[ ] operator is used to extract a continuous function by maintaining continuity across the \( 0 \leftrightarrow 2\pi \) cyclical angular transitions. In order to compare the skeletal curvature of two segments in a scale and rotation-invariant manner, the orientation function \( \Theta_j \) along segment \( SS_j \) must be normalized to unit length and its first-order derivative along \( SS_j \) computed. A set of \((q + 1)\) polynomial coefficients \( \{Q_j\} \) is used to encode the local orientation variations along \( SS_j \). From it, a smaller set of \( q \) polynomial coefficients \( \{dQ_j\} \) representing the scale and rotation-invariant skeletal curvature is derived. They are given by

\[
\{Q_j^k\}_{k=1}^{q+1} = \text{polyfit}\left[\{\frac{1}{N_j}\}_{i=1}^{N_j},\{\Theta_j(i)\}_{i=1}^{N_j}\right]^q \quad \text{and} \quad \{dQ_j^k\}_{k=1}^{q} = \{(i-1)Q_j^{i-2}\}_{i=2}^{q}
\]

(17)

Since the skeletal curvature information is normalized to unit length, the curvature error associated with two segments \( SS_i \) and \( SS_j \) can be computed by subtracting these two polynomials and accumulating the area under the resulting difference polynomial, as shown in Fig. 15. However, in order to apply curvature saliency weighting as described in section 4.3, the area under this difference polynomial is accumulated using the weighting given by the limb width profile along the segment, which incidentally, is also normalised to unit length. Since both the skeletal curvature and limb width profiles are described by the set of \( q \) and \( r \) polynomial coefficients given by \( \{dQ\} \) in (17) and \( \{R\} \) in
(14), the operation described above can be applied directly and efficiently to these coefficients and the saliency-weighted curvature error is given by

\[ d_C(SS_i, SS_j) = \left| \sum_{k=1}^{q+r-1} \left( \frac{1}{k} \{dQ_i^k - dQ_j^k\} \ast \{R_i^k\}_{k=1} \right) \right| \]  

(18)

where \( \ast \) is the convolution operator that is used to implement the multiplication of two polynomials and this results in a new polynomial with \((q+r-1)\) coefficients. The limb width profile used when comparing segment \(SS_i \Rightarrow SS_j\) is that set of polynomial coefficients given by \(\{R_i\}\) and \(SS_i\) is a secondary skeletal segment from the query shape. The selection of polynomial order \(q\) and \(r\) are discussed in more detail in Appendix B.

![Diagram](image)

Fig. 15. The computed skeletal curvature error between skeletons is multiplied by the limb-width variation along the query shape’s skeleton in order to incorporate curvature saliency weighting.

The part distance between two secondary skeletons \(SS_i\) and \(SS_j\) is the weighted combination of these three errors and is given by

\[ d_p(SS_i, SS_j) = \beta_h d_H(SS_i, SS_j) + \beta_c d_C(SS_i, SS_j) + \beta_r d_R(SS_i, SS_j) \]  

(19)

The weights \(\beta_h\), \(\beta_c\) and \(\beta_r\) are used to adjust the contribution of each part descriptor to the overall part distance. Only the histogram error has been normalized to the range of [0,1], the curvature and limb width errors are not. In this sense, the relative values of \(\beta_h\), \(\beta_c\) and \(\beta_r\) do not necessarily indicate their proportional contribution. Suitable values of part descriptors weights were obtained empirically and in most of the experiments presented, the values of \(\beta_h=1.5\), \(\beta_c=0.7\) and \(\beta_r=1.2\) were used.
5.2 Global Distance

In computing the global constraints, a shape $A$ is viewed as being described by an over-complete set of secondary skeleton $\{SS_i^A\}_{i=1}^{NA}$. Each of the $NA$ skeletal parts has two attributes relevant to the computation of global distance. As seen in Fig. 14, they are the 2D coordinate $p_i(1)$ and the local orientation $\theta_i(1)$ at the source node, which results in two types of global relationship between parts, namely global spatial distance and global angular distance. In order to compute these distances, a relatively robust global reference point in the shape has to be identified. It is proposed that a source node (i.e. one least affected by the local deformation and occlusion) from an appropriately selected secondary skeleton be used as a reference point. However, this requires the selection one of $NA$ possible source nodes in shape $A$. This selection process should be a dynamic process that is dependent on the shape $B$ that is being compared. In short, the reference nodes in shapes $A$ and $B$ are selected such that the overall global distance difference between the two shapes is minimized. This selection process is detailed in Appendix A. For now, let us consider that the source node pairing that yields the minimum global distance between shape $A$ and $B$ is given by $p_u^A(1)$ and $p_v^B(1)$, respectively.

5.2.1 Global Spatial Distance

Normalized source node-based spatial distance – The set of global spatial relationships describing the normalized distances between each source node in shapes $A$ and $B$ to their respective reference source nodes (see Fig. 16a) are given by

$$\{TS_i^A\}_{i=1}^{NA} = \left\{ \frac{\|p_u^A(1) - p_i^A(1)\|}{N_{norm}^A} \right\}_{i=1}^{NA}$$

and

$$\{TS_j^B\}_{j=1}^{NB} = \left\{ \frac{\|p_v^B(1) - p_j^B(1)\|}{N_{norm}^B} \right\}_{j=1}^{NB}$$

(20)

where $\|\| \|$ gives the Euclidean distance between two points and the scale-invariant normalization factor $N_{norm}$ is the square root of each respective shape’s total area. The global source node-based spatial distance between two secondary skeletal segments in shapes $A$ and $B$ is then given by

$$g_{ts}(SS_i^A, SS_j^B) = TS_i^A - TS_j^B$$

(21)
This global spatial relationship between parts cannot disambiguate shapes such as those shown in Fig. 16a. In this case, the difference between limb relationships is one that is angular in nature. As such, additional angle-based constraints should be included in the computation of global distance. Two types of normalized angular measures are proposed for each source node, namely the source node skeletal angle and link angle (see Fig. 16b).

Fig. 16. Global distances. (a) Shapes A and B whose limbs differ in their angular but not in their spatial global relationships. (b) The global angular distances used consist of the source node skeletal angle $\theta_i(1)$ and link angle $\psi_i$ that are measured relative to a selected global reference node. They are rotation-normalized by the skeletal angle $\theta_u(1)$ at the global reference node.

5.2.2 Global Angular Distances

Normalized source node skeletal angle – this angular measure is obtained directly from the computed local orientation $\theta_i(1)$ at the source node of each secondary skeleton $SS_i$ (see Fig 16b). In order to ensure rotation invariance, all angular measures are normalized by subtracting the value of the local skeletal orientation $\theta_u(1)$ at the reference source node $p_u(1)$. The set of normalized source node skeletal angles for shape A is therefore given by

$$\{ AS^A_{\theta_i} \}_{i=1}^{NA} = \{ posAngle[\theta_i(1) - \theta_u(1)] \}_{i=1}^{NA} \in [0,2\pi)$$

(22)

where $\theta_u(1)$ is the local skeletal orientation at the reference source node $p_u(1)$. The $posAngle[\cdot]$ operator ensures that the computed angle is represented by a positive value between 0 to $2\pi$. The normalized global source node skeletal angular distance between two secondary skeletons in shapes A and B is the smaller of the two possible angular differences and is given by
\[ g_{as}(SS_i^A, SS_j^B) = \min \left[ \frac{|AS_i^A - AS_j^B|}{\pi}, 2 - \frac{|AS_i^A - AS_j^B|}{\pi} \right] \in [0,1] \tag{23} \]

where || gives the absolute value and the division by \( \pi \) normalises the angular distances such that it has a comparable range to that of the global spatial distance.

**Normalized source node link angle** – this angular measure is computed from the orientation of the line linking a source node to the reference node (see Fig 16b). Given that the reference source node in shape \( A \) has been determined to be \( p_u^A(1) = [x_u^A(1), y_u^A(1)] \), the source node link angle of skeletal segment \( SS_i \) is given by

\[
\psi_i^A = \arctan \left( \frac{y_u^A(1) - y_i^A(1)}{x_u^A(1) - x_i^A(1)} \right) \in [0,2\pi) \tag{24} \]

As in (22), rotation invariance is achieved by subtracting all link angles \( \psi_i^A \) with \( \theta_u^A(1) \) and this results in a set of normalized source node link angles for a shape \( A \) that is given by \( \{AL_i^A\}_{i=1}^{NA} \in [0,2\pi) \). The normalized global source node link angular distance between two secondary skeleton in shapes \( A \) and \( B \) is the smaller of the two possible angular differences and is given by

\[
g_{al}(SS_i^A, SS_j^B) = \min \left[ \frac{|AL_i^A - AL_j^B|}{\pi}, 2 - \frac{|AL_i^A - AL_j^B|}{\pi} \right] \in [0,1] \tag{25} \]

In summary, the global distance between two secondary skeletons in shapes \( A \) and \( B \) is derived by the weighted combination of spatial and angular distances given by

\[
g_d(SS_i^A, SS_j^B) = \gamma_1 g_{as}(SS_i^A, SS_j^B) + \gamma_2 g_{as}(SS_i^A, SS_j^B) + \gamma_3 g_{al}(SS_i^A, SS_j^B) \tag{26} \]

Equal weighting for each of the three different distance measures were used in the experiments presented in section 7.

### 6. Matching Parts and Shapes

Consider the problem of matching query shape \( A \) to shape \( B \), one of the many shapes in a shape database. The shape distance \( s_d(A, B) \) gives the minimal saliency-weighted matching cost that would map all skeletal parts in shape \( A \) to those in shape \( B \). In computing the shape distance \( s_d(A, B) \), each shape \( A \) and \( B \) is viewed as being described
by an over-complete set of secondary skeletal segments \( \{ SS^A_i \}_{i=1}^{NA} \) and \( \{ SS^B_j \}_{j=1}^{NB} \) respectively. As seen in Fig. 17, an over-complete cost matrix \( OCM(i, j) \) of size \( NA \times NB \) is created by computing the sum of the part and global distances for every combination of \( (SS^A_i, SS^B_j) \) and is given by

\[
OCM(i, j) = d_p(SS^A_i, SS^B_j) + g_d(SS^A_i, SS^B_j)
\]  

(27)

where \( d_p() \) and \( g_d() \) are the part and global distances given in (19) and (26) respectively. \( NA \) and \( NB \) are the total number of secondary skeletons in shapes A and B respectively.

Fig. 17. The creation of the over-complete cost matrix \( OCM(i, j) \) and the cost matrix \( CM(i, j) \) from the set of secondary skeletal segments of two shapes A and B

In order to match every skeletal part in shape A to a corresponding part in shape B in a unique and non-repetitive manner, the over-complete matrix \( OCM(i, j) \) is reduced to its basic cost matrix \( CM(i, j) \), as shown in Fig. 17. To achieve this, the minimum visual saliency-weighted matching cost associated with each of the \( NA \) secondary skeletal segments of shape A is computed using

\[
\{ MA^A_i \}_{i=1}^{NA} = \left( \min_{1 \leq j \leq NB} [OCM(i, j)] \right) \times G(SS^A_i)
\]  

(28)
where the part saliency $G()$ is given in (8). Then, for each of the $NM^A$ merged segments in the set \( \{SM^A_m\}_{m=1}^{NM^A} \) of shape $A$, the lowest cost decomposition option of the $2^{NM-1}$ possible combination of secondary skeletons is selected using the matching costs in (28). After all $NM^A$ merged segments have been analyzed, shape $A$ is now described by a basic subset of the original over-complete set of $NA$ secondary skeletal segments given by

$$ \{SR^A_i\}_{i=1}^{nA} \subseteq \{SS^A_i\}_{i=1}^{NA} \quad \text{and} \quad SR^A_i = SS^A_{k(i)} \quad (29) $$

where the set of indices given by $\{k(i)\}_{i=1}^{nA}$ defines $nA$ non-overlapping secondary skeletons selected as the ideal decomposition of shape $A$ with respect to shape $B$. This subset is used to influence the selection of the corresponding basic non-overlapping subset of secondary skeletons for the shape $B$. The minimum saliency-weighted matching cost associated with each of the $NB$ secondary skeletons of shape $B$ is obtained over a reduced over-complete cost matrix using

$$ \{MB_j\}_{j=1}^{NB} = \left( \min_{1 \leq i \leq nA} \left[ OCM(k(i), j) \right] \right) \times G(SS^B_j) \quad (30) $$

Like shape $A$, each of the $NM^B$ merged segment in shape $B$ given by the set \( \{SM^B_m\}_{m=1}^{NM^B} \) is analyzed to determine which decomposition option gives the lowest total saliency-weighted cost given in (30). Shape $B$ is now described by a non-overlapping subset of the original over-complete set of $NB$ secondary skeletal segments as is given by the set

$$ \{SR^B_j\}_{j=1}^{nB} \subseteq \{SS^B_j\}_{j=1}^{NB} \quad \text{where} \quad nB \leq NB \quad (31) $$

With both shapes $A$ and $B$ reduced to their respective basic subsets, a reduced cost matrix $CM(i,j)$ of size $nA \times nB$ is obtained as shown in Fig 17. Since more salient parts in the query shape $A$ should carry a higher matching cost penalty, the cost matrix $CM(i,j)$ is weighted by its respective part’s saliency measure $G()$ given in (8) to produced a saliency-weighted cost matrix $WM(i,j)$ given by

$$ WM(i, j) = CM(i, j) \times G(SR^A_i) \quad (32) $$

With the matrix $WM(i,j)$, the shape distance $s_d(A,B)$ can be obtained by minimizing the total cost of matching subjected to the constraint that the matching be one-to-one. A
square cost matrix results in the classic square assignment problem that can be solved in $O(N^3)$ time by the Hungarian algorithm [3], [8]. The Hungarian algorithm produces a list of skeletal pair matches $(SR^A_i, SR^B_j)$ that results in the lowest overall matching cost given by $s_h(A,B)$. However, the matrix $WM(i,j)$ may not be square. Even though there are extensions of the Hungarian algorithm for rectangular matrices [7], the approach chosen here was to be padded the non-square matrices with dummy costs. The addition of dummy costs serves to penalize a query shape that has additional parts, especially if these additional parts are visually salient. Part saliency weighting is applied to the dummy costs to moderate its influence on the overall cost of matching two shapes when the query shape $A$ has more parts than shape $B$ (see Fig. 18b). The way the dummy costs are employed is dependent on two possible scenarios. When there are less parts in query shape $A$ than in shape $B$ (i.e. $n_A < n_B$), only a subset of parts in shape $B$ will have matches. Cost matrix $WM(i,j)$ is padded with a large dummy cost of say $DC_{A<B} = 10$ (see Fig 18a), which can be subsequently removed to yield a shape distance defined by

$$s_d(A,B) = s_h(A,B) - [(n_B - n_A) \times DC_{A<B}]$$  \hfill (33)

Fig. 18. The method of producing a square weighted cost matrix by adding dummy costs depends on the relative part count between query shape $A$ and shape $B$. (a) When shape $A$ has less parts than $B$. (b) When shape $A$ has more parts than $B$.

When there are more parts in query shape $A$ than in shape $B$ (i.e. $n_A > n_B$) this means some parts in shape $A$ must be matched to dummy parts in shape $B$. Dummy cost equivalent to the average $CM(i,j)$ values of say $DC_{A>B} = 1.5$ is padded to the cost matrix $CM(i,j)$ to make it a square matrix as shown in Fig 18b. If both shapes $A$ and $B$ are generally alike, the extra skeletal parts in $A$ are typically low saliency textural skeletons
(see Fig. 2b). Their impact on the overall matching cost can be reduced by applying saliency weighting to this padded cost matrix using (32) to give $WM'(i,j)$ shown in Fig 18b. The Hungarian algorithm is then applied to the square matrix $WM'(i,j)$ and the shape distance is simply

$$s_d(A, B) = s_h(A, B)$$  \hspace{1cm} (34)

The shape distance between two shapes is not a metric distance, which means $s_d(A, B) \neq s_d(B, A)$. This characteristic is useful as it supports partial matching. On the other hand, this produces difficulty in classifying similar shapes that have some parts missing (see Fig. 19a). In order to overcome such problems, the *bi-directional shape distance* $s_b(A, B)$ is proposed and is given by

$$s_b(A, B) = \frac{1}{2}[s_d(A, B) + s_d(B, A)]$$  \hspace{1cm} (35)

Bi-directional shape distance ensures that a good match is only obtained when the shape distances obtained matching shape $A \rightarrow B$ and $B \rightarrow A$ are both low. It was found that this distance better reflects the similarity between the shapes in the databases used (see Fig. 19b) and was employed in all the shape classification experiments presented.

![Fig. 19.](image)

(a) The shape distance $s_d(A, B)$ does not produce desirable results when the query shape has more parts than other shapes in the same category. (b) Bi-directional shape distance $s_b(A, B)$ provides a better similarity measure under these circumstances.
7. Experimental Results

7.1 Silhouette Classification

The proposed part-based shape analysis (PBSA) strategies implemented within the MGVF skeleton framework was tested on the 99-shape and 216-shape silhouette databases used by Sebastian et al. in [26] (available online at [21]). This test set was used because of the availability of comparative results. As seen in Fig. 20a, the 99-shape database is a difficult data set to classify due to the presence of large within-class variations (four-footed mammal, greeble), limb articulation (man, hand) and occlusions (hare, plane). The first set of results using the 99-shapes database was obtained by matching each shape in the database with all others. Ordered recognition rates in percentage are (100, 100, 100, 100, 100, 100, 100, 99, 91, 88). In other words, the top 7 best matches are correctly classified for all 9 shape categories. This performance is better than Sebastian et. al’s shock graph edit-distance (SGED), which was reported as (100, 100, 100, 99, 99, 99, 97, 96, 95, 87) in [26]. Fig. 20b presents these same comparative results in a precision-recall diagram similar to that used in [26]. Comparison with the contour-based shape context on Belongie et al. in [3] is not entirely fair but it does show the need for an adaptive part-based decomposition strategy when dealing with shape databases featuring significant occlusion and limb articulation.

![The 99-shape database](image)

**Fig. 20.** (a) The 99-shape database and the resulting (b) precision-recall diagram showing the superior shape-based retrieval performance of the MGVF-PBSA algorithm compared to the shock graph edit-distance and the shape context results replicated from [26].
Selected retrieval results from the 99-shape database are shown in Fig. 21. Notice the global inter-part spatial and angular relationships used by MGVF-PBSA (see Fig. 21a) helped reduce mismatches between the dog and human shapes, which are evident in the SGED results (see Fig. 21b). SGED employed only local part descriptors.

Fig. 21. Representative retrieval results using the 99-shapes database. (a) 12 closest matches using MGVF-PBSA, ordered by their bi-directional shape distance and (b) that using shock graph edit distance, ordered by their normalized edit distance. Reproduced from Sebastian et al. results in [26]. Erroneous within-category matches are shaded ( ).

The second set of MGVF-PBSA results using the larger 216-shape database yielded recognition rates of (100, 100, 100, 100, 100, 99, 98, 97, 95, 87, 74). This performance is again better than that of the shock graph edit-distance reported in [26], which only had 100% error-free retrieval for the top 3 best matches, with recognition rates of (100, 100, 100, 99, 97, 99, 96, 96, 95, 91, 80). The more accurate retrieval performance of the MGVF-PBSA is also reflected in the precision-recall graph in Fig. 22b. It performed far better than shape context and is marginally better than SGED. One or two shapes in several categories of the 216-shape database did pose problems for both the MGVF-PBSA and SGED as highlighted in the representative comparative retrieval results shown in Fig. 23. For example, MGVF-PBSA had some problems with the bird and elephant categories (see Fig. 23a). As highlighted in section 7.3, it is not easy to determine a universal weighting between part and global distances that will simultaneously handle large shape variations within a class and part-related distortions due to extreme limb articulation and occlusion. In the case of SGED, the stingray category (see Fig. 23b) was more problematic since curvature saliency weighting was not incorporated.
Fig. 22. (a) The 216-shape database and the resulting (b) precision-recall graph demonstrating the comparative retrieval quality of the MGVF-PBSA, shock graph edit-distance and shape context. For comparison purposes, the MGVF-PBSA result was incorporated into the precision-recall graph presented in [26].

Fig. 23. Representative retrieval results using the 216-shape database. (a) 13 closest matches using the MGVF-PBSA, ordered by their bi-directional shape distance and (b) that using the shock graph edit distance. Reproduced from Sebastian et al. [26].

7.2 Visual Saliency Weighting

Part saliency weighting – This experiment explores the effectiveness of employing the disparity and length based visual saliency measures given in (8) to weigh the computed part matching cost. Fig. 24a shows a rapid fall-off in retrieval accuracy when skeletal segments were given equal weightage during the part distance computation regardless of
its formation stability and length. This result highlights the need to explicitly incorporate ligature-sensitive information and skeletal length in medial-based shape analysis algorithms. Mismatches observed in Fig. 24c suggest that the longer and more stable structural backbone of the greeble shape (which is straight) should be given more weightage during part-based matching so that it can be distinguished from very similar hare shapes (whose backbone is curved). Especially in the presence of large within-class variations in the form of different-shaped ‘ears’ and ‘limbs’.

Fig. 24. (a) Precision-recall diagram showing the positive influence of using visual saliency weighting during shape matching. Sample retrieval results for the greeble category obtained (b) with and (c) without using part saliency weighting. For each query shape (topmost), the 10 best shape matches from the 99-shape database are shown ordered from #1 (best) to #10. Incorrect classifications are shown shaded.

Curvature saliency weighting - Sample retrieval results in Fig. 25 highlights the importance of using the limb-width profile to weigh the skeletal curvature errors computed along the skeleton. Notice in Fig. 25b, the large curvature variations in the elephants’ trunks and the rays’ tails increased misclassifications in the absence of curvature saliency weighting.

Fig. 25. Sample retrieval results from the 216-shape database [26] (a) with and (b) without curvature saliency weighting.
7.3 Incorporating Global Distance

Shape retrieval accuracy degraded significantly when inter-part spatial and angular relationships (global distance) were not taken into account during shape matching (see Fig. 26a). The nature of the mismatches is clearly observed in Fig. 26c when only local part descriptors were employed. Many hand shapes were incorrectly retrieved as their fingers found good local part matches with the mammals’ legs despite obvious geometrical relationship differences. This result supports the use of global distance especially if large within-class variations exist in the data set.

![Precision-recall diagram showing the positive influence of using global inter-part geometry during shape matching.](image)

Fig. 26. (a) Precision-recall diagram showing the positive influence of using global inter-part geometry during shape matching. Sample retrieval results for the 4-footed mammal category are shown (b) with and (c) without using global distance information.

However, excessive global distance emphasis in not always favorable, especially if the objects within the same category have articulating parts (see Fig 27a). As global distance emphasis was reduced, perfect retrieval results were then obtained (see Fig. 27b).

![Excessive global distance weightage penalizes retrieval of object with articulating parts.](image)

Fig. 27. Excessive global distance weightage penalizes retrieval of object with articulating parts. (a) Ordered classification results obtained for a data set of articulating objects and another when (b) the global distance contribution was reduced by a factor of 0.3 relative to the part distance.
7.4 Computational Cost

The computational costs associated with creating and querying a 25-shape database (used in [3]) were examined. Timing results in Table 1 show that the time taken to create the MGVF-PBSA database is significantly longer than that taken to create the 25 shape contexts [3]. Unlike the shape context (a global descriptor), the MGVF-PBSA breaks a shape down into parts and their associated multiple descriptors have to be individually computed. In typical database applications, the computation of shape descriptors is only done once when a new shape first enters the database. Of more concern is the time taken to query the database. From Table 1, it can be seen that MGVF-PBSA’s query time is at least 10 times faster than those of shape context’s. Though both the matching algorithms use the $O(N^3)$ Hungarian algorithm to solve a square assignment problem, the dimension of the MGVF-PBSA’s cost matrix is typically between the sizes of $8 \times 8$ to $20 \times 20$ (for the 25-shape database) compared to the shape context’s $100 \times 100$ (for $N=100$). Other significant computation done in MGVF-PBSA is the matching of the three $n$-bin orientation histograms per secondary skeleton. Fortunately, the computation of $\chi^2$ distances between histograms is a fast $O(n)$ operation. The timing comparison with shock graph edit-distance (SGED) was not done as public domain codes were not available. Computational complexity analysis presented in [26] suggest that for a non-heuristic algorithm that compares two tree $T_1$ and $T_2$ requires a total time of $O(n_1^3 n_2^3)$, where $n_1$ and $n_2$ are the number of nodes in $T_1$ and $T_2$ respectively. Assuming that the typical number of nodes in SGED trees is similar to that of the MGVF-PBSA part decomposition, the order of complexity in the SGED shape matching operation is significantly higher than the $O(N^3)$ of the MGVF-PBSA. This however, does not take into account other computation cost (e.g. histogram comparison) required to derived an over-complete cost matrix $OCM(i,j)$.

<table>
<thead>
<tr>
<th>Shape Descriptor</th>
<th>Computational Cost (seconds)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Create database</td>
</tr>
<tr>
<td>MGVF-PBSA</td>
<td>390.80</td>
</tr>
<tr>
<td>Shape Context</td>
<td>42.38</td>
</tr>
</tbody>
</table>

Table 1 – Comparative computational costs for extracting/matching the MGVF-PBSA descriptors and shape context [3] (with $N=100$). The times shown were measured on a 1.6 GHz Pentium-4 PC with 255 MB, running Windows 98® and Matlab® version 5.3.
8 Conclusions

Several strategies were proposed to improve the performance of skeleton-based shape matching algorithms. Firstly, the unavoidable presence of ligatures in medial-based shape representation must be explicitly accounted for. It was shown how this can be implemented in both the part matching and part decomposition processes. In the former, as a form of visual saliency measure that can be used to weigh the computed part matching cost. In the case of part decomposition, ligature-sensitive information was incorporated into the merging process of basic skeletal components such that the numerous combinations of these components can be suitably reduced in a visually coherent manner based on the notion of visual continuity along the skeleton. Additionally, the selection of the shape’s part decomposition option is deferred until the shape matching stage. This is done in acknowledgement of the fact that the ideal decomposition of a shape is dependent on the shape it is being compared with. The use of curvature saliency along a skeleton was also suggested as a means to moderate the effect of curvature error when comparing skeletal parts. Additionally, experimental results suggest that shape matching can be made more robust if both local and global inter-part measures are used to compute shape dissimilarity. However, it was observed that the emphasis between local and global distances is somewhat dependent on the nature of the shape matching application and data set. These numerous strategies were implemented and tested within the MGVF skeleton framework. Experimental results showed that incorporation of these suggested strategies do produced significant improvement in shape retrieval accuracy. Application of these strategies on other skeleton-based shape matching algorithms should yield similar benefits, albeit their implementations may be realized differently.

9 References


Appendix A – Finding the Global Reference Nodes

In order to compute the global distance between two shapes \( A \) and \( B \), the global reference node pair given by \( (p_u^A, p_v^B) \) must first be determined as summarized in Fig. 28.

---

**Fig. 28.** Summary of procedure to determine global reference nodes for shapes \( A \) and \( B \).

---

\[ \xi_{gd} = \sum_{i=1}^{NA} \min_{1 \leq j \leq NB} [g_d(SS_i^A, SS_j^B)] + \sum_{j=1}^{NB} \min_{1 \leq i \leq NA} [g_d(SS_i^A, SS_j^B)] \quad (37) \]

\( NA \) and \( NB \) are the total number of secondary skeletons \( (SS_n) \) in shapes \( A \) and \( B \) respectively. And \( g_d(\cdot) \) is the normalized global distance between the two segments given in (26).
Appendix B – Experiments on some MGVF-PBSA Parameters

Several critical parameters of the MGVF-PBSA algorithms were altered to study their influence on the shape classification performance. The 99-shape database was used for this purpose.

![Graph A](image1.png) ![Graph B](image2.png)

**Fig. 29.** The precision-recall graphs for varying MGVF-PBSA parameters. Results from (a) varying the orientation histogram bin size $n$ and (b) varying the $q$ and $r$ polynomial orders for the skeletal curvature and limb width descriptors respectively.

The orientation histogram bin size given by $n$ in equations (9)-(12) have a significant influence on the data storage requirement of the shape descriptor. It can be observed in Fig. 29a that as $n$ increases, the shape classification performance improves. This improvement is due to the increased resolution used to encode the gradient vector field local shape and connectivity characteristics of the skeletal part. However, not much further improvement is obtained going beyond $n = 36$ (see [14]). Another experiment was done on the $q$ and $r$ polynomial orders for the skeletal curvature and limb width descriptors in (17) and (14) respectively. Results in Fig. 29b show that by moving from an order of 3 to 6, only a slight improvement in performance is obtained. Further increase of polynomial order from 6 to 9 did not yield further improvement. A high-order polynomial with a large number of coefficients is undesirable as it translates to a higher data storage requirement for the part descriptors. From these findings, the parameter settings used for the orientation histogram bin size was $n = 36$ and the polynomial orders for describing the skeletal curvature and limb width profile were $q = r = 6$.  

36