Chapter 7

Matching Pairs of Silhouette Sets

7.1 Introduction

This chapter moves on to the next major topic: recognising individual stones from their silhouettes. The key idea is to apply the ET-based pose optimisation described in Chapter 6 to pairs of silhouette sets. If the residual ET error after pose alignment is sufficiently low, then the pair of silhouette sets is classified as a match (i.e., produced by the same stone); otherwise, the pair of silhouette sets is classified as a mismatch (i.e., the two silhouette sets were produced by two different stones).

Recall that the ultimate goal for the recognition component of this thesis, as stated in Chapter 1, is *batch matching*. Batch matching is matching two *batches* of silhouette sets from two unordered runs of the same batch of stones. This chapter investigates the simpler problem of *verification*, i.e., verifying that a pair of silhouette sets was produced by the same stone (a match). The methods developed for verification will be extended in later chapters for the purposes of batch matching.

The proposed alignment-based method achieves its accuracy by approaching the matching problem from the point of view of silhouette consistency, rather than considering the similarity between 3D approximations of the stones computed from each silhouette set. A weakness of using 3D approximations is that the shape of the 3D approximations will vary with stone orientation even in the noise-free case. Chapter 8 considers a matching method that is based on 3D approximations to stone shape. Although less accurate than the method describe in this chapter, it is substantially faster to compute. Chapter 9 will demonstrate how the two methods can be combined to achieve both speed and accuracy for solving the batch matching problem.

Alignment-based matching simply requires applying the ET-based pose optimisation described in Chapter 6. This chapter investigates two modifications to the method: (1) the use of an orthographic projection model, and (2) the use of a measure of inconsistency based on the cone intersection projection (CIP) constraint.

The use of an orthographic projection model is aimed at improving efficiency. The use of a measure of inconsistency based on the CIP constraint is aimed at improving accuracy.

Experiments are carried out on image sets of stones captured using the six-camera setup and the two-mirror setup. The experiments demonstrate how the residual ET error across pairs of silhouette sets separate match cases from mismatch cases. The improvement in running time efficiency is quantified for the use of an orthographic projection model. The effect of using CIP-based measures of inconsistency is investigated using downsampled real data and synthetic data for different camera configurations. (Downsampling is required to create overlap between the match and mismatch distributions.) Synthetic data are used to investigate the behaviour of ET-based matching on different camera configurations and at different levels of image resolution.

7.2 Related Work

There is a wealth of literature on recognising silhouettes from a fixed viewpoint, a 2D recognition problem. Since this thesis considers stones that are arbitrarily oriented with respect to the cameras, these approaches are not relevant. The computer vision literature describes several approaches to silhouette-based matching from variable viewpoints. The principal difference between the problems addressed by these methods and the problem addressed in this work is that single silhouettes are used for matching, whereas here silhouette sets are used. Several approaches are outlined below and their relevance to this work is explained.

Jacobs et al. [65] consider the problem of recognising an object from a single silhouette. Their method is related to the approach described in this chapter in that recognition is attempted without 3D reconstruction, only a small number of views is used, and silhouette consistency is used to determine matches. However, the authors limit themselves to the case in which the camera translates and rotates about a known axis that is parallel to the image. Outer tangents are used to determine consistency using an approach based on linear programming.

Lazebnik et al. [77] describe a method for recognising objects from single silhouettes by storing multiple silhouettes of objects in a database. A geometrical approach is taken, where a match is considered to correspond to consistent epipolar geometry. The method achieves its discriminatory power by considering all epipolar tangents rather than only the outer epipolar tangents. For epipolar tangents to aid discrimination, the tangencies must be visible across different viewpoints. Since most stone silhouettes do not have epipolar tangents (other than the outer epipolar tangents) that are visible across multiple viewpoints, such an approach is not feasible.

7.3 An Orthographic Model for Computing ET Error

In this section, a method for computing ET error using an orthographic imaging model is described. An orthographic model offers several advantages over the perspective model:

1. An analytical expression can be computed for the Jacobian matrix that is used by the Levenberg-Marquardt routine for pose optimisation. This has the potential to speed up the computation. Without an analytical expression, the Jacobian is estimated using a forward difference method [58]. This method requires one extra evaluation of the cost function for each dimension of the pose parameter vector. Since seven parameters are used to describe the pose (a quaternion and a 3D translation vector), each evaluation of the Jacobian requires seven extra evaluations of the cost function. If the analytical expression can be computed faster than this, then the matching process can be completed in less time.

2. In the perspective case, it is possible that an epipole may lie within a silhouette. If this occurs, there will be no outer epipolar tangencies. The use of a perspective model requires the additional overhead of identifying these cases, and introduces the additional complexity of differing numbers of reprojection errors corresponding to different poses. In the orthographic case the epipoles are always at infinity, and thus correspond to directions [58]. Each silhouette image of a stone will always yield two outer epipolar tangencies with respect to the epipole.

3. Tangencies can be computed more efficiently using an orthographic imaging model. This is described in Section 3.5.4. The gain in efficiency is because directions of epipolar tangencies correspond exactly to the direction of the epipole for an orthographic imaging model. Tangencies can therefore be unambiguously determined using the edge-angle data structure. Unlike the perspective case, no check is required to confirm that the vertex is a tangency.

4. Unlike in the case of a perspective model, for an orthographic model residual errors (i.e., distances from tangencies to projected epipolar tangents) computed in one image of a pair are identical to residual errors computed in the other image of the pair. This means that residual errors need only be computed in one image for each pair.

The orthographic model is computed separately for each silhouette view and is based on the perspective model for each of the cameras, which is determined with the once-off camera calibration procedure. The orthographic model is thus a very close approximation to the full perspective model in the vicinity of the stone.

The original polygonal boundary is used to create an approximation to the orthographic projection that would be observed from an orthographic camera that shares a viewing direction with the perspective camera. Whereas the original polygonal boundary is specified in pixel units, the orthographic approximation

is specified in world units (millimetres). To approximate an orthographic vertex $(p_x, p_y)^\mathsf{T}$ from the image coordinates $(u,v)^\mathsf{T}$ of a vertex of the original polygonal boundary, the following equation is used:

$$\begin{pmatrix} p_x \\ p_y \end{pmatrix} = \left[\begin{pmatrix} u \\ v \end{pmatrix} - \begin{pmatrix} u_0 \\ v_0 \end{pmatrix} \right] \frac{z_{\text{cen}}}{f}, \tag{7.1}$$

where z_{cen} is the depth of the centroid of the VEMH. The assumption that is implicit in Equation 7.1 is that the depth of the VEMH centroid closely approximates the depth of the rim points.

Consider computing the ET error across two silhouette sets: Set A and Set B. This requires reprojection errors to be computed across each silhouette from Set A paired with each silhouette from Set B (as described in Section 3.5.2).

Reprojection errors are computed for each pair of silhouette views. The relative pose of two views (one from Set A and one from Set B) is described by a rotation R followed by a translation $\mathbf{t}$ that transform points in the reference frame of View 2 into the reference frame of View 1:

$$\mathsf{R} = \begin{pmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{pmatrix}, \tag{7.2}$$

$$\mathbf{t} = \begin{pmatrix} t_x \\ t_y \\ t_z \end{pmatrix}. \tag{7.3}$$

The last rows of R and $\mathbf{t}$ are not needed, since orthographic projections are used and any point may thus slide arbitrarily along the Z-axis of either camera. However, in practice R and $\mathbf{t}$ are computed as in Equations 7.2 and 7.3 since they are computed directly from 4×4 rigid body transform matrices that are used to describe silhouette poses in different reference frames. The 4×4 rigid transform matrix $M_{C2 \to C1}$ that transforms from Camera 2's reference frame to Camera 1's reference frame is computed as follows:

$$M_{C2 \to C1} = \begin{pmatrix} \mathsf{R} & \mathbf{t} \\ \mathbf{0}^\mathsf{T} & 1 \end{pmatrix} = (M_{WA \to C1})(M_{\text{WB} \to \text{WA}})(M_{C2 \to WB}). \tag{7.4}$$

The two silhouette sets are Set A and Set B. Camera 1 is from Set A and Camera 2 is from Set B. The matrix $M_{WA \to C1}$ describes the rigid transform from Set A's world reference frame to Camera 1's reference frame. The matrix $M_{\text{WB} \to \text{WA}}$ describes the rigid transform that attempts to align Set B's world reference frame with Set A's world reference frame. The matrix $M_{C2 \to WB}$ describes the rigid body transform from Camera 2's reference frame to Set B's world reference frame (usually computed as $M^{-1}_{WB \to C2}$). The matrices $M_{WA \to C1}$ and $M_{WB \to C2}$ are computed using a once-off camera calibration procedure; the candidate pose is represented by the matrix $M_{\text{WB} \to \text{WA}}$. To compute R and $\mathbf{t}$ for a pair of silhouettes and a candidate pose, a 4×4 matrix representation of the candidate pose must be formed, and then Equation 7.4 is used. Note that the candidate pose describes the pose between Set A and Set B, so the pose between specific views within Sets A and B must be derived from both the candidate pose and the relative poses within a set.

The reprojection error is the distance from an outer epipolar tangency to the epipolar line corresponding to the tangency in the opposite view. In the noise-free case the distance will be zero, since the tangencies are two views of the same 3D point (a frontier point). In order to determine the tangencies, it is first necessary to determine the epipolar directions. The tangencies can then be located, and the reprojection errors can be computed.

The epipolar direction $\mathbf{e}_{12}$ is the projection of the viewing direction of View 2 onto the image plane of View 1. This is illustrated in the example shown in Figure 7.1a.

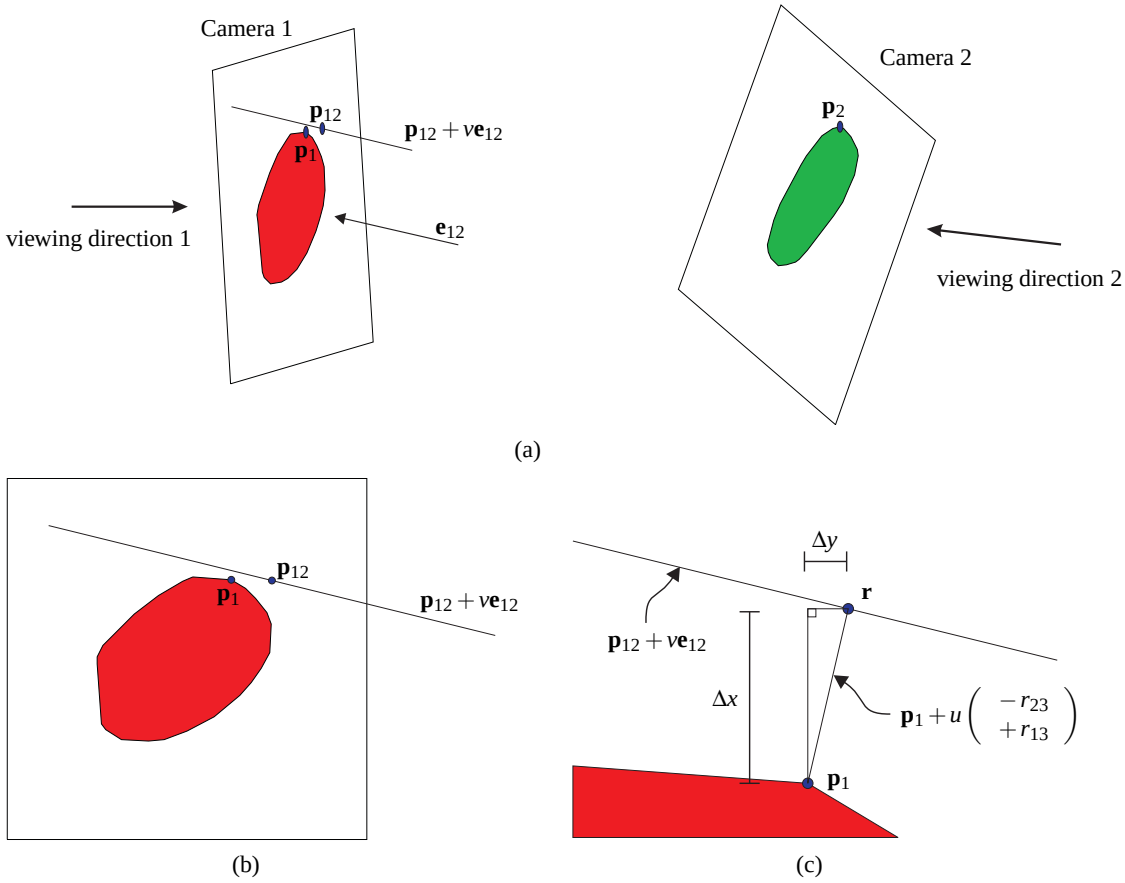


Figure 7.1: The epipolar geometry relating two orthographic views of a stone: (a) shows the two silhouette views, (b) shows the image plane of the first view, and (c) shows a closeup of (b) in the vicinity of the outer tangency under consideration.

The viewing direction $\mathbf{d}$ of a camera in its own reference frame is

$$\mathbf{d} = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}, \tag{7.5}$$

since by convention the camera is modelled to point along the z-axis. In the reference frame of Camera 1, the viewing direction of Camera 2 is

$$\mathbf{d}_{12} = \mathsf{R}\mathbf{d} = \begin{pmatrix} r_{13} \\ r_{23} \\ r_{33} \end{pmatrix}. \tag{7.6}$$

Since $\mathbf{d}_{12}$ is a direction, it is unaffected by $\mathbf{t}$. The epipolar direction $\mathbf{e}_{12}$ is obtained by dropping the z-coordinate of $\mathbf{d}_{12}$ to project the vector onto the image plane of Camera 1:

$$\mathbf{e}_{12} = \begin{pmatrix} r_{13} \\ r_{23} \end{pmatrix}. \tag{7.7}$$

Once the epipolar direction has been computed, the tangency vertices are located using the edge-angle data structure as described in Section 3.5.4.

The reprojection errors are computed as the distance from an epipolar tangency to the epipolar line corresponding to the opposite epipolar tangency.

To compute the epipolar line that the projection of $\mathbf{p}_2$ is constrained to lie on (given the relative pose between Views 1 and 2), a point on the line is considered. For simplicity, the z-coordinate of $\mathbf{p}_2$ in the reference frame of Camera 2 is set to zero so that

$$\mathbf{p}_2 = \begin{pmatrix} p_{2x} \\ p_{2y} \\ 0 \end{pmatrix}. \tag{7.8}$$

The projection $\mathbf{p}_{12}$ is then given by

$$\mathbf{p}_{12} = \begin{pmatrix} r_{11}p_{2x} + r_{12}p_{2y} + t_x \\ r_{21}p_{2x} + r_{22}p_{2y} + t_y \end{pmatrix}. \tag{7.9}$$

The epipolar line $\mathbf{p}_{12} + v\mathbf{e}_{12}$ meets the line that passes through the epipolar tangency $\mathbf{p}_1$ and is perpendicular to the epipolar direction at

$$\mathbf{r} = \mathbf{p}_{12} + v_0\mathbf{e}_{12} = \mathbf{p}_1 + u_0 \begin{pmatrix} -r_{23} \\ +r_{13} \end{pmatrix}. \tag{7.10}$$

(See Figure 7.1.) Solving for u_0 gives

$$u_0 = \frac{r_{23}(p_{1x} - r_{11}p_{2x} - r_{12}p_{2y} - t_x) - r_{13}(p_{1y} - r_{21}p_{2x} - r_{22}p_{2y} - t_y)}{r_{23}^2 + r_{13}^2}. \tag{7.11}$$

Reprojection errors Δx and Δy are then given by

$$\begin{pmatrix} \Delta x \\ \Delta y \end{pmatrix} = u_0 \begin{pmatrix} -r_{23} \\ +r_{13} \end{pmatrix}. \tag{7.12}$$

The above equations provide an efficient means for computing the ET error in the case of an orthographic imaging model. The residual values should only be computed in one image plane of the pair. This is because the distance values in both images are equal to the distance between two specifications of the epipolar tangent plane: one specified by the camera centres and the epipolar tangency in View 1, and the other specified by the camera centres and the epipolar tangency in View 2.

An important advantage of the equations laid out above is that an analytical expression can be derived for the Jacobian matrix that is used by the Levenberg-Marquardt method that is used for pose optimisation. The derivation of this Jacobian matrix is described in Appendix B.

7.4 Error Formulations Based on the CIP Constraint

The ET constraint is a weaker constraint than the CIP constraint (as described in Section 3.5). The ET constraint specifies a necessary, but insufficient condition for consistency. An error or degree of inconsistency derived using the ET constraint has been chosen for pose optimisation, since it is efficient to evaluate. This is important since pose estimation is an iterative procedure that requires the error to be evaluated for different parameters over many iterations.

Under noisy conditions, a mismatch pair of silhouette sets may have an ET error that is sufficiently low that the pair is misclassified as a match. However, since the CIP constraint is stronger than the ET constraint (it specifies both a sufficient and a necessary condition for consistency), a measure of inconsistency based on the CIP constraint may not result in a misclassification. This approach makes use of the ET error for pose optimisation, but uses a once-off evaluation of a CIP-based error for match verification.

This section presents three measures of inconsistency that are based on the CIP constraint: Boyer error, convex CIP error, and nonconvex CIP error. Boyer error is based on Boyer's silhouette calibration ratio [14, 15], whereas convex CIP error and nonconvex CIP error are novel formulations. The three measures of inconsistency will be compared with ET error in terms of match verification accuracy.

7.4.1 Boyer Error

Boyer's method [14, 15] considers the rays corresponding to each silhouette point in the silhouette regions that are not covered by the CIP. The error associated with each ray is determined by computing the 3D point on the ray that is consistent with the largest number of the silhouettes in the set. The error contributed by this ray is proportional to the number of silhouettes that are inconsistent with the 3D point (i.e., silhouette viewpoints in which the 3D point does not project into the silhouette).

Since there is a continuum of rays corresponding to any finite image region, silhouettes are pixelated in order to compute the Boyer error. The ray passing through the pixel centre is considered for each pixel. Higher resolution pixelations will therefore lead to more accurate approximations of the Boyer error.

The ray corresponding to each foreground pixel in each silhouette view is considered in turn. Each of the remaining silhouettes defines a (possibly empty) interval on the ray for which the silhouette is consistent with the ray. This interval is computed by projecting the ray onto the silhouette. The projected ray-silhouette intersection is a line segment. This line segment is projected back onto the 3D ray to obtain the interval. A 3D point on the ray that lies within the maximum possible number of intervals is considered next. If the maximum number of intervals is equal to the number of silhouettes, then the pixel is covered by the CIP, and does not contribute to the error since it is consistent with all silhouettes. Otherwise, the pixel contributes an error of $ka/(A(m-1))$, where k is the number of the $m-1$ remaining silhouettes that are consistent, a is the pixel area, and A is the total area of foreground regions in the silhouette set.

Boyer's method differs from the other methods (such as the approach of Hernández [39], and the formulations that are presented in Sections 7.4.2 and 7.4.3) in that the viewing ray corresponding to each point within a silhouette is not simply classified as consistent or inconsistent. Rather, a degree of inconsistency is obtained for each viewing ray. This is done by using the *number* of consistent silhouettes for each viewing ray. Another approach would be to take into account some measure of *distance* from consistency for each silhouette-ray pair. This was not implemented because it is inefficient to compute, making it impractical to apply to a large number of silhouette sets.

7.4.2 Convex CIP Error

The convex CIP error is an attempt to create an error formulation that is fast to compute by limiting the input to convex silhouettes. The approach achieves efficiency by providing a closed form solution that is computed directly from the input polygons (as opposed to the other methods which require rasterisation). This comes at the cost of discarding information in the concave regions of the original polygonal boundaries. This information may potentially aid discrimination between matches and mismatches.

The convex hulls of the silhouette boundaries are used as input. This approach works because the silhouettes of the 3D convex hull of a stone are the 2D convex hulls of the silhouettes of the stone: if a silhouette set is consistent, then a silhouette set formed from the 2D convex hulls of the silhouettes in this set will be consistent too.

The convex CIP error method integrates the distance (possibly raised to some power of n) from the silhouette boundary to the CIP over the silhouette boundary for all silhouettes in the set. Silhouette regions that are not overlapped by the CIP can contribute error according to how far they are removed from the CIP (by squaring or cubing the distance for instance). This is a potential advantage over the Boyer error, since uncovered

regions that are far from the boundary are more likely to be caused by a mismatch than by segmentation or calibration error.

Using convex silhouettes provides several advantages:

1. The cone intersection can be efficiently computed.
2. The boundaries of the CIP can be easily and efficiently computed.
3. A boundary-based error formulation can be used, since convex boundaries cannot have fractal-like perimeters that make nonconvex boundaries highly sensitive to resolution.

The cone intersection of convex cones can be efficiently computed as a halfspace intersection. Each cone face represents a halfspace (the halfspace on the cone side of the plane passing through the face). The halfspace intersection is computed using a dual space formulation which allows a convex hull algorithm to be used. Efficient convex hull algorithms exist ($O(m \log m)$ for m cone faces).

The boundaries of the CIPs can now be obtained by computing the 2D convex hulls of the projected vertices of the cone intersection. Note that computing the boundary of a projection of a nonconvex polyhedron is a far more elaborate procedure.

To integrate the distance (raised to the nth power) to the CIP around the silhouette boundary, the boundary is traversed to identify triangular and trapezial* regions (see Figure 7.2). The triangles consist of portions of

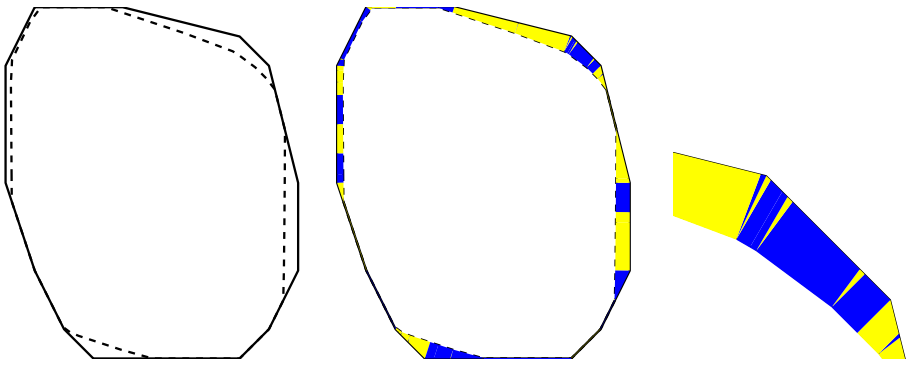


Figure 7.2: Computing convex CIP error: (a) the silhouette boundary (solid line) and CIP boundary (dashed line), (b) the triangles and trapezia that must be considered when computing the error (with alternating shading to aid visualisation), and (c) a closeup of the top right of (b).

the silhouette boundary whose closest point is the same point on the CIP. The trapezia are made up of the remaining portions of the boundary. The trapezia and triangles are computed by determining whether the closest point on the CIP polygon is along an edge (for trapezia) or at a vertex (for triangles).

*Confusingly, what is referred to as a trapezium in British English is a trapezoid in American English, and vice versa. Here, British English is used: a trapezium is a quadrilateral with a pair of parallel sides.

The error component E_{tri} contributed by each of the triangles is computed from the three side lengths s, t, and u (as illustrated in Figure 7.3) as follows (the side of length u lies on the silhouette boundary polygon):

$$E_{\text{tri}} = \int_{-w}^{0} (x^2 + f^2)^{n/2} dx + \int_{0}^{v} (x^2 + f^2)^{n/2} dx, \tag{7.13}$$

where

$$\theta = \text{acos}((t^2 - s^2 - u^2)/(-2su)) \tag{7.14}$$

$$f = s \sin\theta \tag{7.15}$$

$$v = s \cos\theta \tag{7.16}$$

$$w = u - v. \tag{7.17}$$

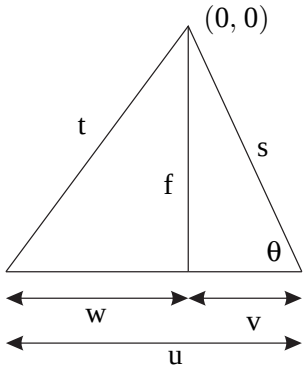


Figure 7.3: Diagram for computing the error contribution of triangles.

Equation 7.13 was successfully evaluated for values of $n = 1, 2, 3$, and 4 using the Matlab Symbolic Toolbox. In this work, the exponent of $n = 2$ was used. The solution to the integral for $n = 2$ is

$$E_{\text{tri}} = \frac{1}{(18ut^2 + 18us^2 - 6u^3)}. \tag{7.18}$$

Similarly, the error component E_{trap} contributed by each of the trapezia is computed from the four side lengths p, q, r, and m (as illustrated in Figure 7.4) as follows (the side of length m lies on the silhouette boundary polygon):

$$E_{\text{trap}} = \int_{0}^{m} \left[\frac{x}{m} p + \left(1 - \frac{x}{m}\right) r\right]^n dx = m(p^{n+1} - r^{n+1})/((p - r)(n + 1)). \tag{7.19}$$

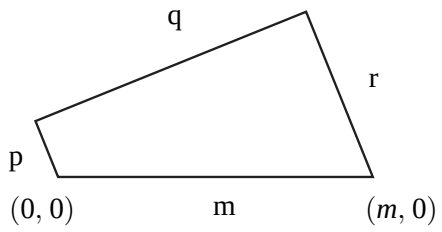


Figure 7.4: Diagram for computing the error contribution of trapezia.

7.4.3 Nonconvex CIP Error

Nonconvex CIP error works with the silhouettes directly rather than their convex hulls. This means that saddle-shaped regions on a stone that are imaged as concave portions of the silhouette boundary can potentially be used to discriminate between matches and mismatches. For instance, if two stones have very similar convex hulls, yet differ in shape because of saddle-shaped regions, the nonconvex CIP error may be able to correctly classify a mismatch that would be misclassified if the convex CIP error were used. The nonconvex CIP error is therefore only likely to show a substantial improvement over convex CIP error for matching applications in which the stones have substantial variation in the shapes of saddle-shaped regions of their surfaces.

Unlike the convex CIP error, area-based integrals are used, since a nonconvex boundary can be highly sensitive to resolution. The integral of the distance to the CIP (raised to the nth power for some n) from all points inside the silhouette, but outside the CIP, forms the nonconvex CIP error.

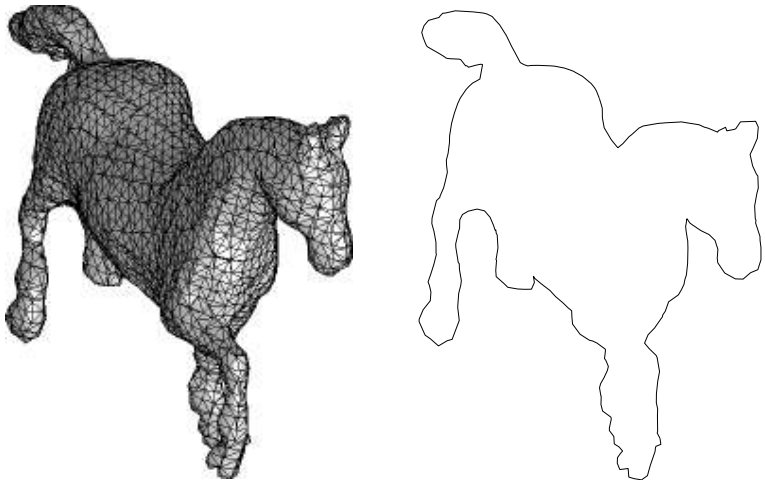
A visual hull algorithm is used to compute the cone intersection[†]. Although Matusik et al. [91] describe an efficient algorithm, it is not used here because it produces a so-called polygon soup output that does not provide connectivity information for the faces. This makes computing an exact projection of the cone intersection impossible, and creating an approximate quantised projection of the cone intersection by rendering is slow. Instead, a marching tetrahedron-based approximation to the visual hull was created using a C++ implementation[‡] of Bloomenthal's implicit surface polygonizer [8]. The routine outputs the visual hull surface as a triangular mesh. An alternative algorithm is the exact method described by Franco and Boyer [48].

The CIP outline is computed exactly from the triangular mesh representation. To do this, it is necessary to identify the contour generator edges on the triangular mesh with respect to the viewpoint. Only edges formed by a pair of faces in which exactly one face is towards the viewpoint are candidates. The bottommost vertex of the projected edges is guaranteed to lie on the outline, and may be used as a starting point. From the starting point, the algorithm moves from vertex to vertex. Care must be taken in selecting the correct edge, as multiple candidate edges may share a common vertex. Edge projections may also be crossed by edges from

[†]In the case of a match, the cone intersection is the visual hull of the object and associated viewpoints. However, to avoid abuse of terminology, the term cone intersection is preferred since in the case of a mismatch set the visual hull is not a meaningful concept.

[‡]The polygonizer library was provided by J. Andreas Bærentzen (Technical University Denmark).

a different part of the triangular mesh. These crossings cause a discontinuity in the contour generator (but not in its projection), and introduce new vertices. Each other candidate edge must therefore be checked when traversing each edge. This makes the algorithm of $O(m^2)$ complexity for m candidate edges. The algorithm terminates upon returning to the starting point. An example is shown in Figure 7.5.



(a) (b)

Figure 7.5: Example showing (a) a triangular mesh, and (b) the outline of a projection of a triangular mesh computed using the algorithm described above.

For each silhouette, the CIP and the silhouette are rendered to create rasterised versions. A distance transform is used to assign values to pixels in the silhouette according to their distance from the silhouette outline. Pixels not covered by the CIP contribute error of the distance value. Error values are summed for all pixels over all silhouettes. The approach of computing the exact projection of a triangular mesh described above allows the error to be computed reasonably efficiently. Note that by counting the number of pixels that are not covered by the CIP, one obtains an estimate of the area of non-overlap. This measure was considered by Hernández et al. [60], who created a perimeter-based method to approximate the area of non-overlap.

7.5 Experiments

7.5.1 Empirical Match and Mismatch Distributions

Pose optimisation was applied to match and mismatch pairs of silhouette sets formed from the data set of five runs of 246 garnet stones captured with the six-camera setup. The five silhouette sets of each stone (from the five runs) can be paired in $5!/(2!(5-2)!) = 10$ different ways. This means that there are $10 \times 246 = 2460$ match pairs. The $5 \times 246 = 1230$ silhouette sets can be paired in $1230!/(2!(1230-2)!) = 755\,835$ ways of which $755\,835 - 1230 = 754\,605$ are mismatch pairs. However, only mismatches across different runs were considered, as this still provides a large number of mismatch pairs, whilst simplifying the selection of

the pairs. This means that $10 \times (246 \times 246 - 246) = 602\,700$ mismatch pairs were considered. Note that although there is a large number of mismatch cases, they are not statistically independent samples. This is because the same stone is used in multiple pairs. Such dependence is common in the evaluation of biometrics recognition systems such as fingerprint verification. Bolle et al. [10] have developed the subsets bootstrap to account for this dependence when estimating uncertainty associated with accuracy performance statistics derived from such data.

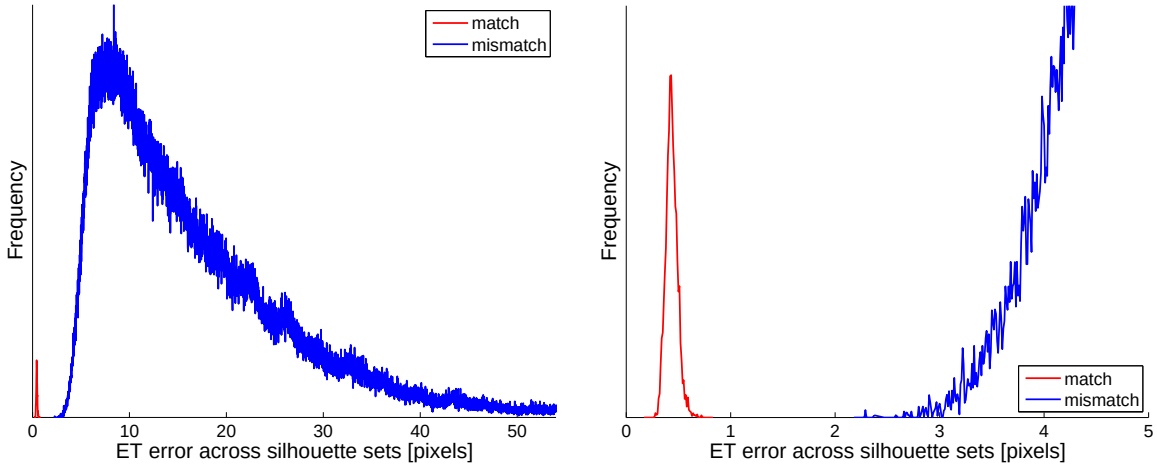
Pose optimisation based on the orthographic projection model was used, as the garnets are small with respect to the distances to the camera. One hundred starting points were used for each silhouette set pair. The first four starting points were based on alignment of the principal axes of the two VEMHs, and the remaining orientation components were uniform random rotations. The experiment required several days of running time on a 3.2 GHz Pentium 4 machine.

Figure 7.6 shows the distributions of ET error values across silhouette sets after pose optimisation for matches and mismatches. It is important to notice that the match and mismatch distributions do not overlap: the smallest mismatch ET error is 2.26 pixels, whereas the largest match ET error is 0.72 pixels. This means that there is a range of ET error thresholds (0.72–2.26 pixels) that will separate, without misclassifications, match pairs from mismatch pairs for the pairs considered in this data set.

Figure 7.7 illustrates the pairs corresponding to the highest match error (i.e., the match that comes closest to being misclassified as a mismatch), and the lowest mismatch error (i.e., the mismatch that comes closest to being misclassified as a match). As can be seen in the figure, the mismatch pair exhibits a higher degree of inconsistency than the match pair both in terms of ET error distances shown in red, and in terms of the degree of non-coverage of the silhouettes by the CIPs.

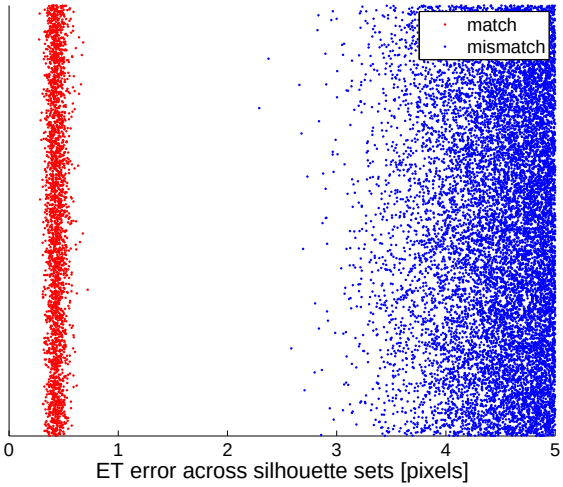
The non-overlap between the match and mismatch distributions, is beneficial to the accuracy of matching based on pose optimisation. However, it makes estimating the long run error rate and comparison between different methods difficult. In Section 7.5.4, ET error is compared with CIP-based error formulations by using reduced resolution images with smaller numbers of views. Reducing the resolution and number of views creates overlap between the match and mismatch distributions, which facilitates comparison of different methods.

The experiment was repeated with the gravel data set captured using the mirror setup. Three runs of five-view silhouette sets of the 220 gravel stones were used. From this data, $3 \times 220 = 660$ match pairs and $(220 \times 220 - 220) \times 3 = 144\,540$ mismatch pairs were formed. Unlike in the case of the garnet data, pose optimisation was applied using the perspective model. This was because some initial experimentation indicated that the camera was sufficiently close to the stones for the orthographic model to be inappropriate. The experiment required more than a week's processing time. The results of the experiment are shown in Figure 7.8. Again, there is a range of ET threshold values that completely separate the match pairs from the mismatch pairs from this data. In the case of the gravel data, the largest ET error across silhouette sets for a match pair is 0.68 pixels, whereas the smallest ET error from a mismatch pair is 3.15 pixels.



(a)

(b)



(c)

Figure 7.6: ET errors across silhouette set pairs after pose optimisation for 2460 match pairs and 602 700 mismatch pairs formed from the six-view garnet data set: (a) shows match and mismatch distributions estimated with a kernel smoothing method (a low variance Gaussian was used to limit over-smoothing); (b) shows a closeup of (a) in the region where the distributions are closest; (c) shows the data points in the region illustrated by (b) (the vertical component, which is random, is a visualisation aid).

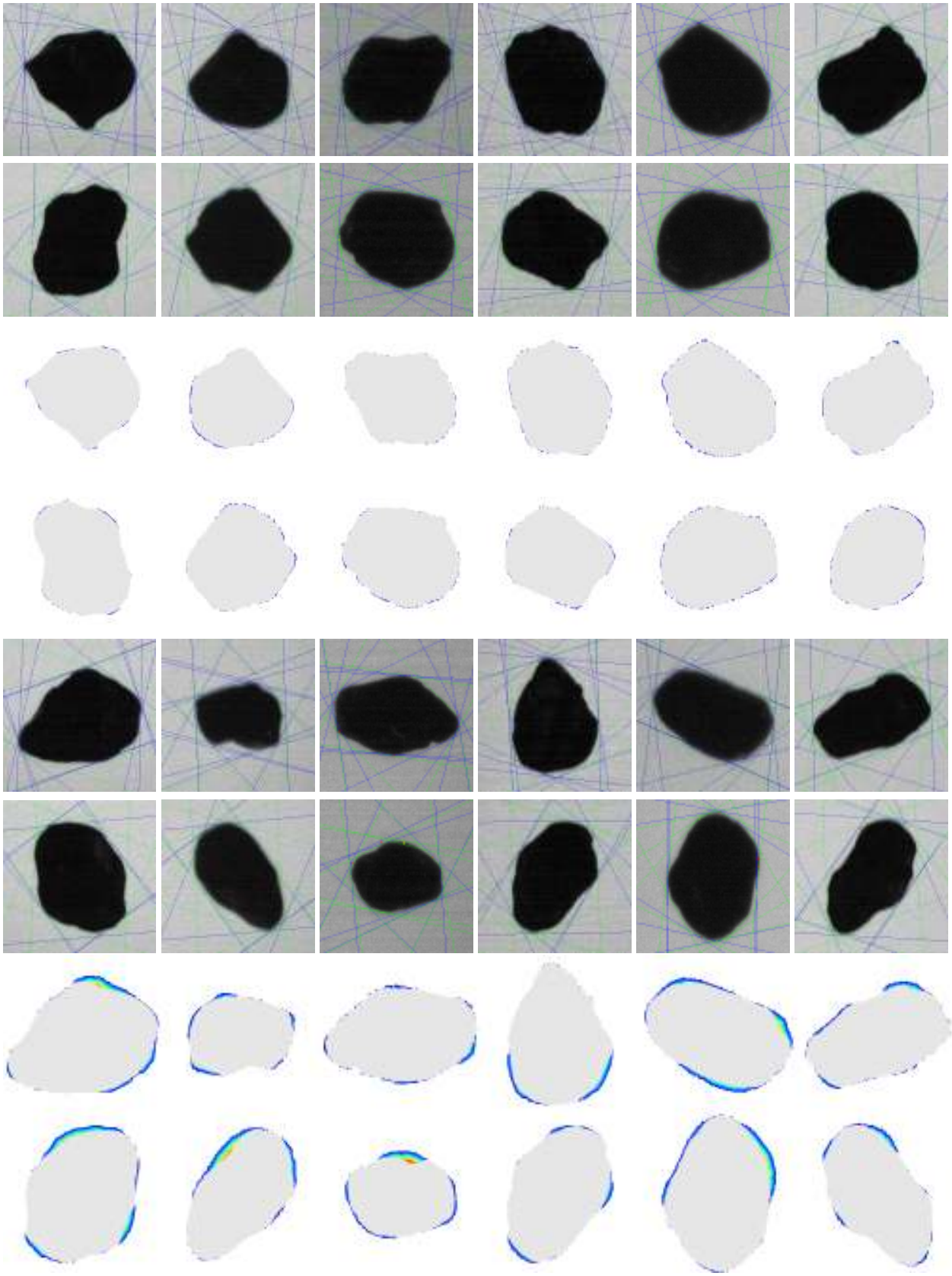


Figure 7.7: Pair of garnet silhouette sets corresponding to the highest ET error of a match *(top half)*, and the lowest ET error of a mismatch *(bottom half)*. For each pair, the top row shows projected epipolar tangents within the silhouette set in green, across silhouette sets in blue, and error distances in red; the bottom row shows silhouettes in colour with 12-view CIPs in grey.

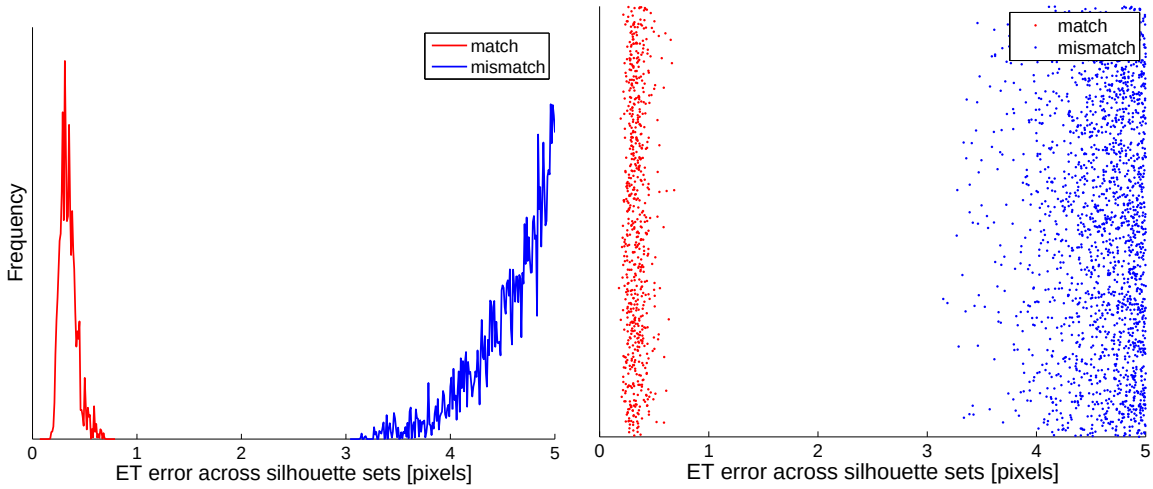


Figure 7.8: ET errors across silhouette set pairs after pose optimisation for 660 match pairs and 144 540 mismatch pairs formed from the five-view gravel data set captured using the mirror setup: (a) shows the region where the distributions are closest; (b) shows the data points in the region illustrated by (a) (the vertical component, which is random, is a visualisation aid).

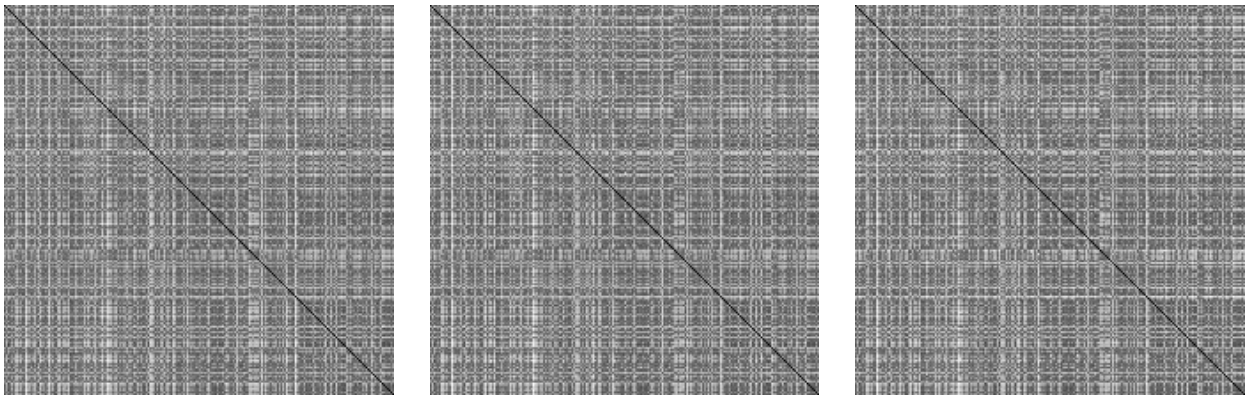


Figure 7.9: Match matrices formed from the gravel data set. Each element corresponds to a silhouette set pair from two runs. Diagonal elements are match pairs and off-diagonal elements are mismatch pairs. The three matrices correspond to the three run pair combinations. Darker regions indicate lower ET error.

Figure 7.9 illustrates the three 220×220 match matrices formed from the gravel data set.

7.5.2 Recognising Stones by Mass

The results observed in the above-mentioned experiments indicate that ET error after pose optimisation provides a potentially accurate way to recognise individual stones. It is interesting to consider another method that one might use to identify or recognise individual stones on different occasions: the stone's mass as measured by an electronic balance.

Figure 7.10 shows ROC (receiver operating characteristic) curves computed using mass difference as measured by an electronic balance as a measure of dissimilarity. Results are shown for the data set of gravel

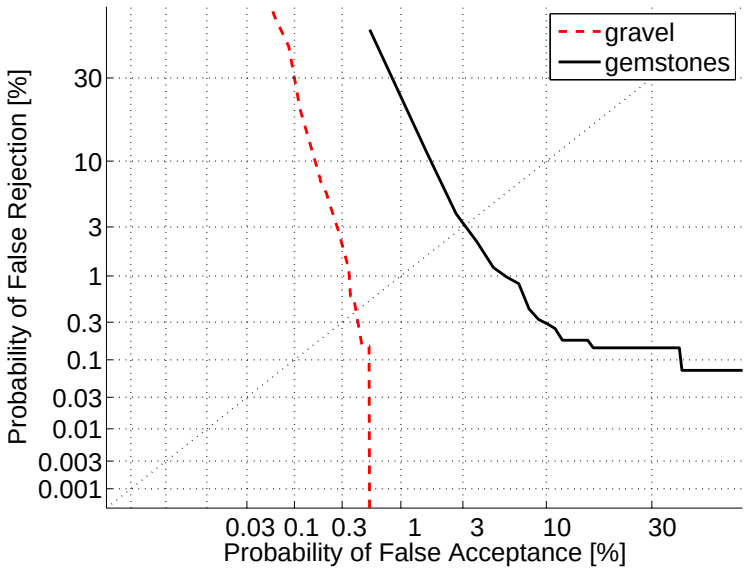


Figure 7.10: ROC curves computed for dissimilarity defined as the difference in mass as measured by an electronic balance. The ROC curves have nonlinear axes to aid visualisation: this is the detection error tradeoff curve format introduced by Martin et al. [89].

and gemstones. Each stone was weighed with an electronic balance. Each gravel stone was weighed on three different days (three runs of mass measurements). Two runs of mass measurements were captured for the gemstone data. This was carried out by staff of the company that provided the gemstones. The ROC curves show the estimated error rates that one would obtain from matching the stones based on a threshold on the difference between mass values measured on two different occasions. The ROC curves are computed by determining all the measured mass differences for matches and for mismatches across runs. The ROC curves indicate that using measured masses cannot be used to provide error-free classification. This is because the variability of measured masses is sufficiently high to create differences in measured mass of the same stone that are in some cases higher than the mass differences between different stones. In the case of the gravel stones, the data indicate that the equal error rate is approximately 0.5%, whereas the equal error rate is approximately 3% for the gemstone data. The gemstone data produce larger errors partially because the mass variability is not as large as for the gravel data: there are many cases where pairs of gemstones have approximately the same mass, and are therefore prone to being misclassified as matches (a false acceptance). The ROC curve for the gemstone data indicates that to attain a low false rejection rate (say $< 0.1\%$), one must tolerate a very high false acceptance rate ($> 30\%$). This is because the mass measurements contain several gross errors (the measurements were possibly incorrectly transcribed by the data capturer). A few large measured mass differences for match pairs result in the necessity of a high tolerance for measured mass differences if one is to ensure that the false reject rate remains low. This will result in a high false acceptance rate as many mismatch pairs will be misclassified as matches. The principal reason for presenting the ROC curves in Figure 7.10 is to demonstrate that identifying stones by individual mass is infeasible for the data sets considered in this thesis.

7.5.3 Running Time Experiments

Section 7.3 presented an algorithm for computing ET error based on an orthographic projection model. This approach was designed to speed up the computation of the ET error. The results presented here quantify the speedup that one obtains using the orthographic model instead of the perspective model.

Table 7.1 shows the mean running time for optimisation from a single starting point for different optimisation types. Results were computed using 246 pairs of six-view silhouette sets of garnets.

imaging model	tangency location	Jacobian computation	mean time
orthographic	lookup	analytical	4.6 ms
orthographic	scan	analytical	11.6 ms
orthographic	lookup	forward difference	8.6 ms
orthographic	scan	forward difference	28.6 ms
perspective	lookup	forward difference	73.6 ms
perspective	scan	forward difference	144.6 ms

Table 7.1: Mean running time per optimisation for various methods: 'lookup' means the edge-angle lookup was used and 'scan' means that each vertex of the polygon was visited to determine tangencies. Times are computed using 6-view silhouette sets of garnets with a stopping criterion requiring an error reduction of no more than 1% reduction of the RMS residual ET error over three Levenberg-Marquardt steps. A 3.2 GHz Pentium 4 machine was used.

The results demonstrate the speedup that is achieved in practice when using the proposed modifications to compute ET error. A speedup of a factor of 30 is achieved over the basic perspective model without tangency lookup. Tangency lookup increases the speed of the perspective-based method by a factor of two. A further speedup of a factor of approximately eight is achieved by switching to an orthographic model. The use of an analytical expression for the Jacobian matrix provides a further speedup of more than a factor of two.

7.5.4 Performance of CIP-Based Error Formulations

Silhouette inconsistency formulations based on the CIP constraint were presented in Section 7.4. To compare the performance of the CIP-based formulations with ET error, the garnet images were downsampled to reduce the silhouette boundary accuracy, and in so doing to create an overlap between the match and mismatch distributions for ET error. Each pixel in 32×32 blocks of pixels was replaced with the mean intensity value of the 32×32 block. This mimics what would be obtained using a camera with lower resolution. Figure 7.11 shows an example of a downsampled silhouette set.

Pose optimisation as described in Section 7.5.1 was applied to matches and mismatches formed from the first two runs of the downsampled garnet data. (Only two runs were used because of the long running time required for these experiments.) The experiment was repeated using different subsets of camera views to investigate the effects of varying the number of cameras.

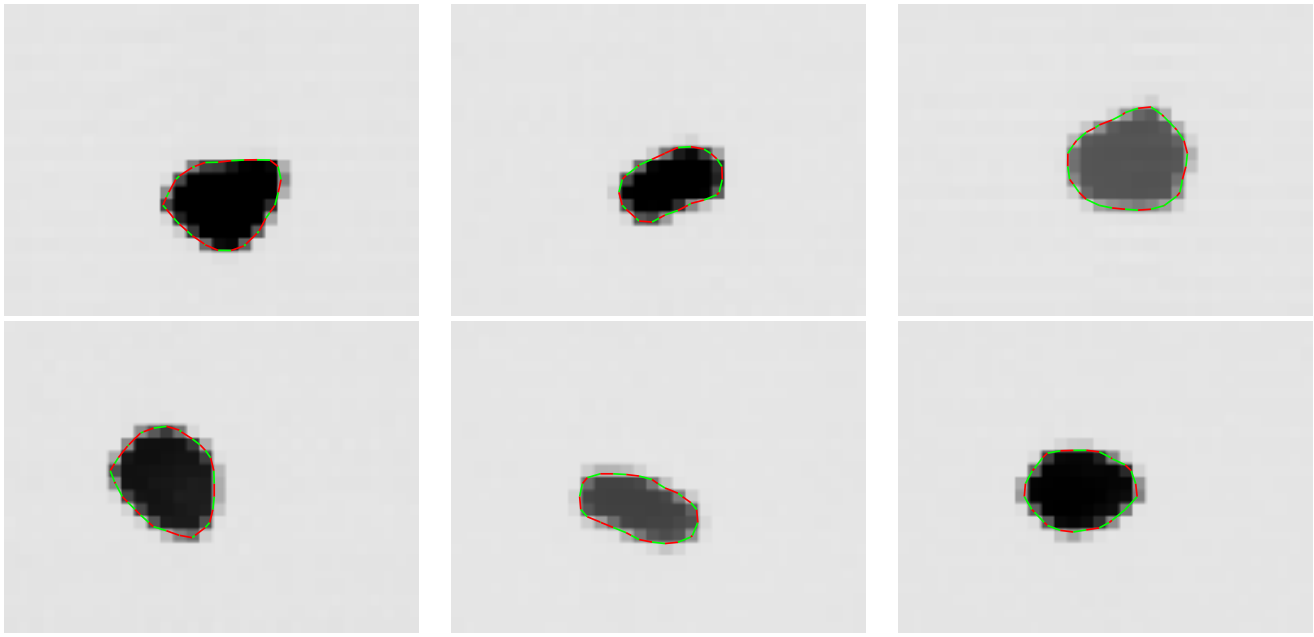


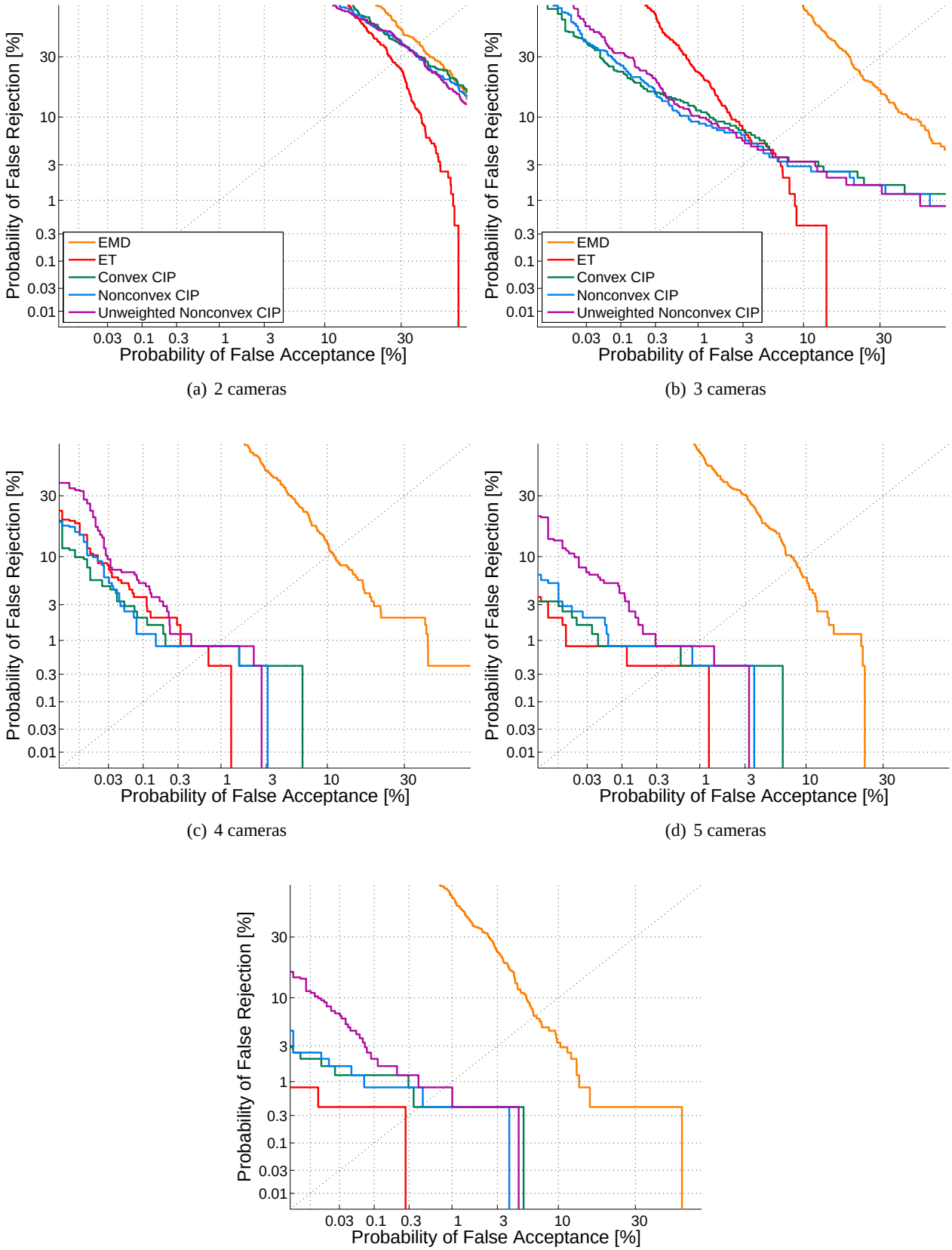
Figure 7.11: An example of a six-view image set of a garnet after 32×32 downsampling. Extracted polygonal boundaries are shown in colour.

Where parameter values needed to be chosen (e.g., silhouette rasterisation resolution for Boyer error, and voxelisation resolution for nonconvex CIP error), the values were chosen so that any attempt to further increase the accuracy would result in negligible improvements. The approach to parameter value selection therefore sacrifices speed in favour of matching accuracy.

Figure 7.12 shows ROC curves computed after ET-based pose optimisation using different CIP-based formulations and ET error. Boyer error was not computed in these experiments because of its prohibitively high running time. The plots show ROC curves based on an additional measure of dissimilarity: earth mover's distance (EMD) between caliper distributions of the VEMH. This method will be described in the next chapter. Unweighted nonconvex CIP refers to nonconvex CIP error computed without the distance transform: the silhouette area not covered by the CIP is used without weighting uncovered regions according to the distance from the boundary.

The plots show that greater accuracy is achieved as the number of cameras is increased, because this increases the number of consistency constraints imposed by the silhouettes within a set. Despite incorporating more information than ET error, Figure 7.12 indicates that in most cases the CIP-based error formulations produce worse ROC curves than the ET error. The plots suggest that the CIP-based methods only outperform ET error (in terms of accuracy) for certain operating points of the ROC curve for the 3-camera case.

The poor performance of the CIP-based methods (particularly for more than three cameras) may be a consequence of the following:



(a) 2 cameras

(b) 3 cameras

(c) 4 cameras

(d) 5 cameras

(e) 6 cameras

Figure 7.12: ROC curves computed using two runs of downsampled images from the garnet data set. Results are shown for different subsets of the six camera views. (See text for further details.)

1. As the number of well-distributed views is increased, the constraints imposed by the epipolar tangents become closer to the constraints imposed by the CIPs. This was pointed out by Hernández [39]. The potential advantage of using a CIP-based measure of inconsistency is therefore diminished as the number of views considered is increased.

2. CIP-based errors require the cone intersection to be computed. The cone intersection is sensitive to noise, since if any silhouette indicates that a 3D region of space is empty it is considered to be empty. This noise sensitivity is analogous to using a maximum rather than a mean of some feature to characterise a class of objects. For instance, one might expect the maximum caliper diameter of a silhouette to be more noise sensitive than the mean caliper diameter.

3. Stones tend not to have deep concavities that are visible from widely disparate views. This limits the potential of nonconvex CIP error to incorporate information that cannot be captured by ET error (or convex CIP error).

Figure 7.12b suggests that CIP-based error can provide superior performance to ET error in at least some situations. Figure 7.13 uses bootstrap replications to illustrate that the observed differences between the ROC curves are not likely to be due to chance alone. When considering the statistical variability of a curve estimated from samples, Efron [35] recommends using bootstrap samples (i.e., repeatedly drawing n samples from the original n samples with replacement) for a "quick and dependable picture of the statistical variability in the original curve." The idea is that the variability of the bootstrap curves approximates the variability that one would obtain if one carried out the same experiment (with different random samples) many times. Since error values are not independent, the 'subsets bootstrap' method of Bolle et al. [10] was used. The method groups error values in an attempt to reduce dependence as much as possible. The plot indicates that the observed superior performance of the CIP-based methods in the upper region persists over twenty bootstrap replications.

The superior performance of CIP-based methods for certain operating points in the three-view case is clearly of little practical significance, and thus far the CIP-based methods appear to be of little use.

In a further attempt to investigate whether CIP-based methods might outperform ET error in certain situations, synthetic data sets were used. The first set was created from refined visual hull models of 100 uncut gemstones that were selected for their degree of nonconvexity. The stones are scaled along their three principal axes so that each stone has unit convex volume and its caliper diameter along the three principal directions are in the ratio $2 : 3 : 4$. Giving the stones the same gross shape ensures that false acceptance errors occur. The synthetic nonconvex stones are illustrated in Appendix C on page 225. Nonconvex stones were created as these have the potential to demonstrate the superiority of nonconvex CIP error over convex CIP error.

A further set of synthetic stone shapes was generated based on the convex hulls of the refined visual hull models of the first 200 garnets. Again, the stones are scaled along their three principal axes so that each

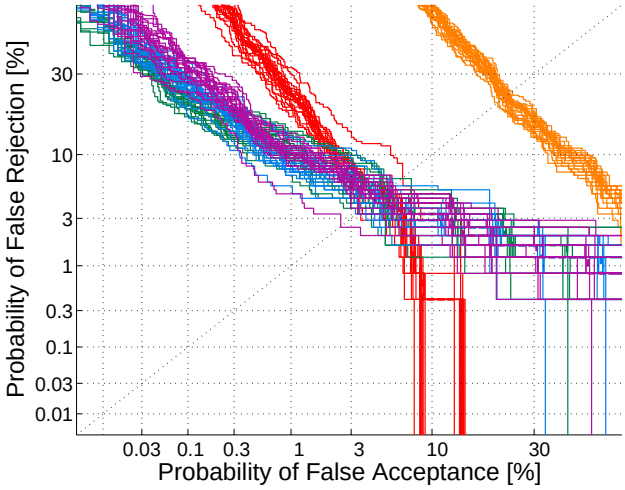


Figure 7.13: Twenty bootstrap replications of the ROC curve shown in Figure 7.12b.

stone has unit convex volume and its caliper diameter along the three principal directions are in the ratio 2 : 3 : 4. These stones are illustrated in Appendix C on page 226.

Synthetic images were generated from the synthetic stones by rasterising the polygonal projections. Various camera configurations were simulated. The camera configurations are based on the optimised frontier point criterion (as described in Chapter 5, and illustrated in Figure 5.5 on page 86).

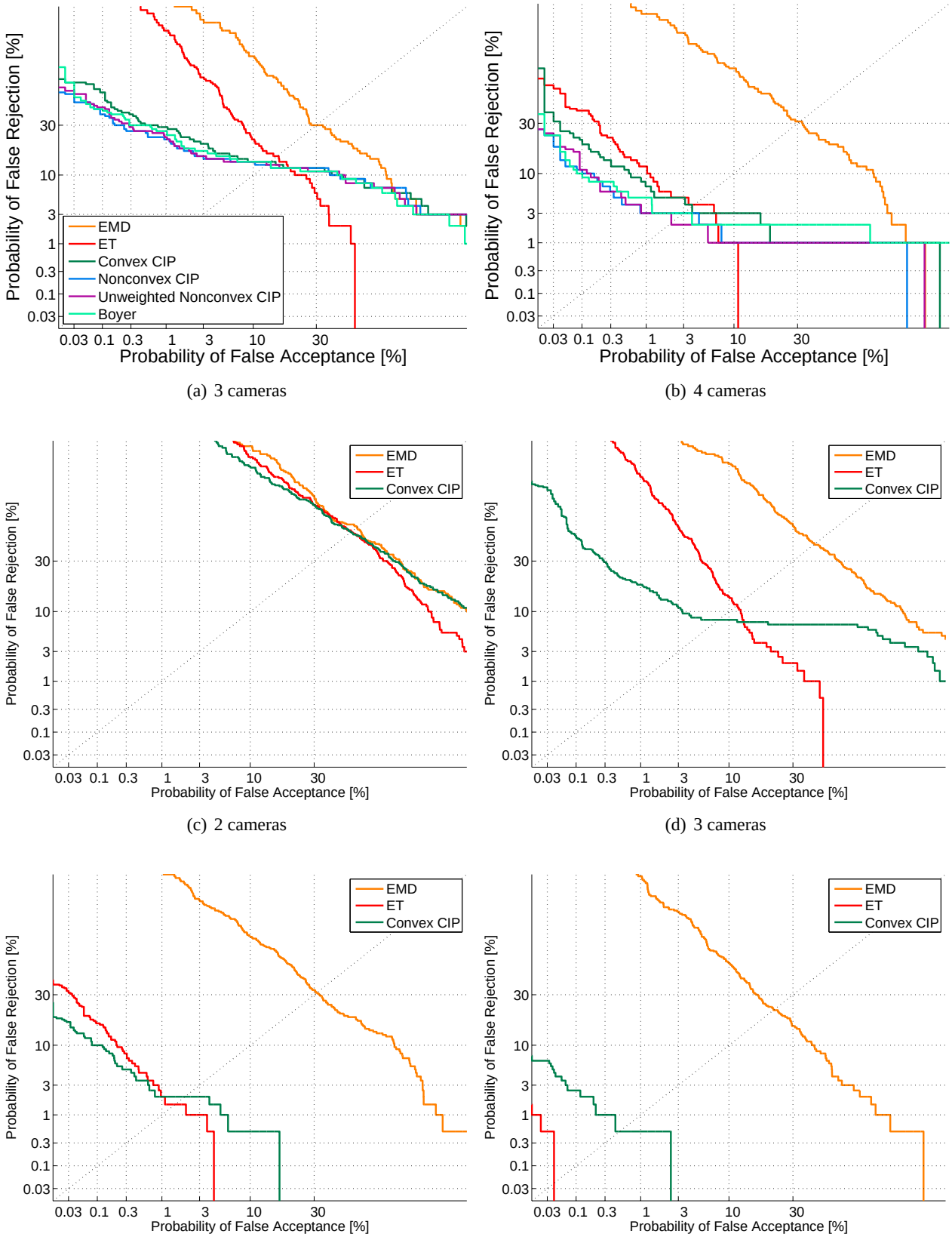
ROC curves computed from the synthetic data are shown in Figure 7.14. Boyer error was computed for the nonconvex stones. The plots show similar behaviour to the real data: CIP-based methods outperform ET error only for small numbers of cameras (fewer than six), and in these cases the outperformance is only for certain operating points on the ROC curve.

7.5.5 Effect of Image Resolution and Camera Configuration

Synthetic data sets were further used to investigate the effects of camera configuration and image resolution on match and mismatch distributions of ET error after pose alignment.

Camera configurations based on the optimisation criteria described in Chapter 5 were used. In addition, various six-camera configurations were investigated to illustrate the importance of the configuration of cameras for a fixed number of cameras. The additional six-camera setups were generated by varying the elevation angle of cameras positioned in a semicircle. Three examples are shown in Figure 7.15.

Figure 7.16 shows the match and mismatch ET error values for different camera configurations and factors of resolution reduction. Many of the camera configurations show a similar trend: as the image resolution



(a) 3 cameras

(b) 4 cameras

(c) 2 cameras

(d) 3 cameras

(e) 4 cameras

(f) 6 cameras

Figure 7.14: ROC curves computed using synthetic data sets: (a)–(b) nonconvex stones, (c)–(f) convex stones.

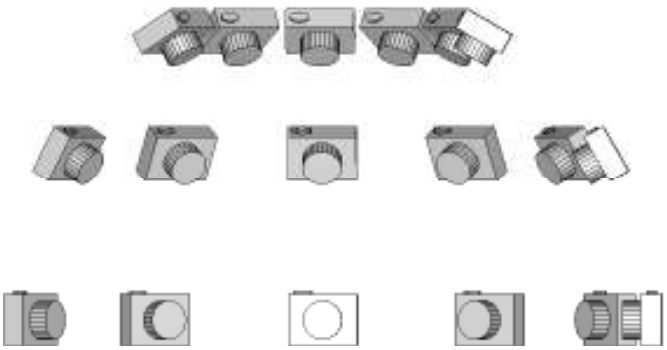


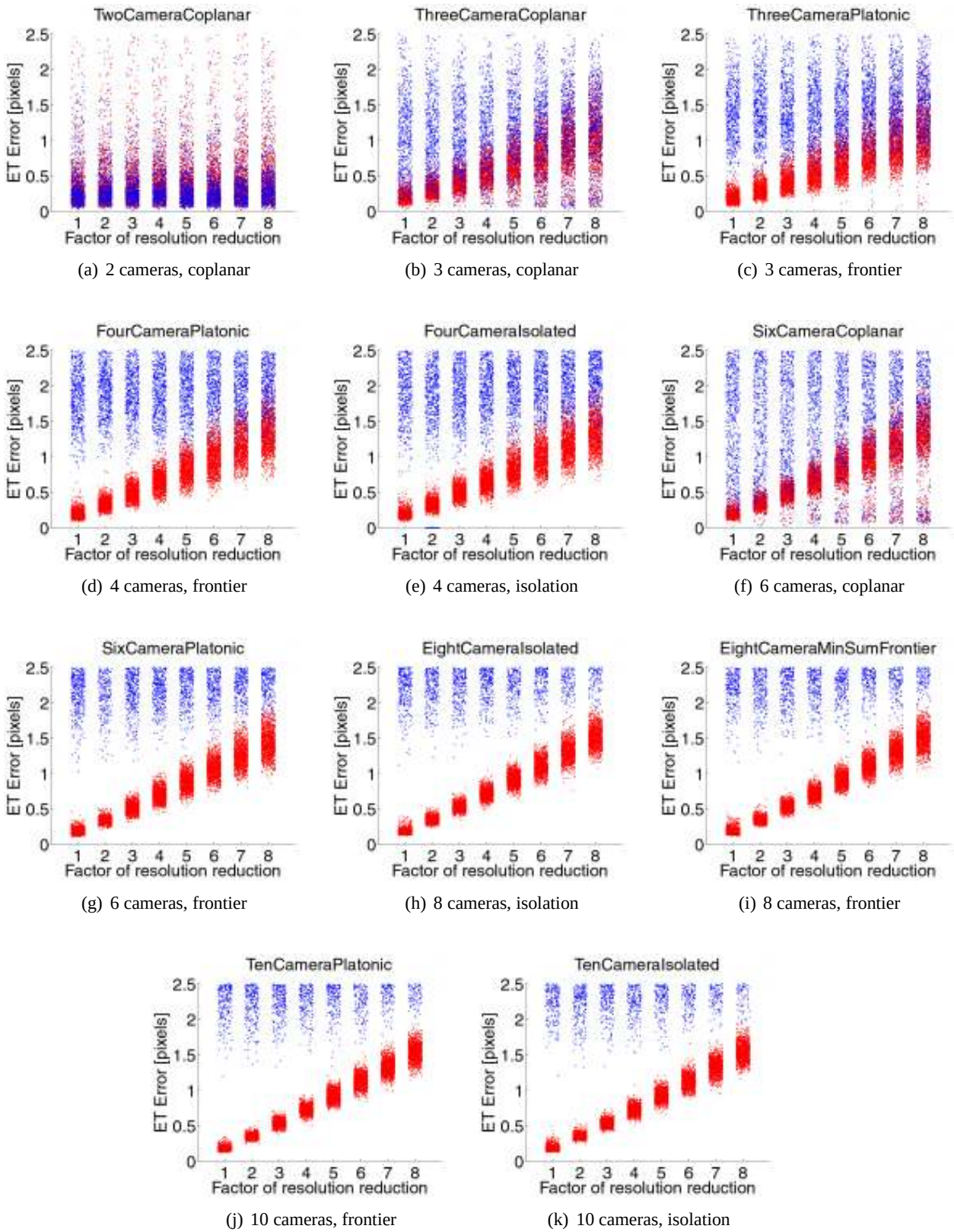
Figure 7.15: Six-camera setups with elevation angles of 60° *(top row)*, 30° *(middle row)*, and 0° *(bottom row)*.

is decreased, the distribution of match errors rises in a predicable fashion, and the distribution of mismatch errors stays in roughly the same position, rising only slightly.

There are, however, some pathological cases: in some cases cameras with coplanar optical axes produce low ET errors for both match and mismatch pairs. This is because the epipolar tangencies occur in approximately the same position for coplanar camera setups. (The extent to which the epipolar tangencies are not exactly coincident is influenced by the degree of perspective distortion: if the cameras are moved back to infinity, the epipolar tangencies will be exactly coincident.) The plots therefore clearly illustrate the undesirability of the coplanar camera configuration for matching.

In the case of well-distributed cameras, camera configurations based on minimising the most isolated viewing direction, and on minimising the sum of distances between frontier points produce similar results. This is because any well-distributed camera setup is likely to produce epipolar tangencies that are well-separated from one another, and whose residual error values are almost independent from one another.

Figure 7.17 illustrates the match and mismatch error values for different six-view configurations. The plots demonstrate that the configuration of cameras is important. The camera configurations correspond to different elevation angles. For low elevation angles, the configuration is close to the coplanar configuration, and poor separation between match and mismatch error values is observed. As the elevation angle is increased the separation improves, and then degrades again as the elevation angle becomes large and the viewing directions converge.



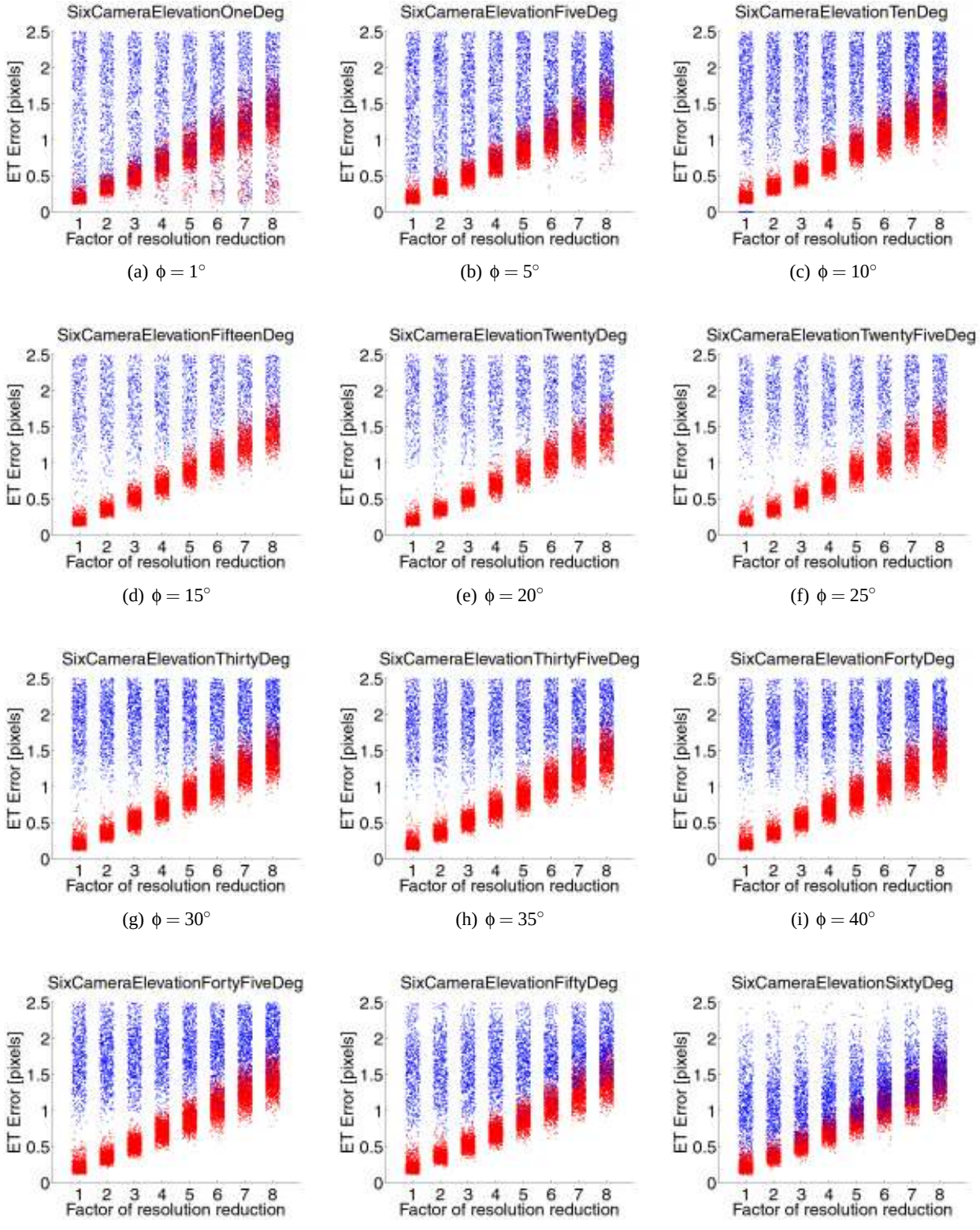
(a) 2 cameras, coplanar (b) 3 cameras, coplanar (c) 3 cameras, frontier

(d) 4 cameras, frontier (e) 4 cameras, isolation (f) 6 cameras, coplanar

(g) 6 cameras, frontier (h) 8 cameras, isolation (i) 8 cameras, frontier

(j) 10 cameras, frontier (k) 10 cameras, isolation

Figure 7.16: Plots of match values (red) and mismatch values (blue) for different camera configurations. Different levels of quantisation noise are shown, corresponding to different degrees of resolution reduction from the original image resolution. Cameras are configured to fulfil the frontier point criterion (frontier), the direction isolation criterion (isolation), or to have coplanar optical axes with even angular distribution about 180° (coplanar).



(a) $\phi = 1^\circ$ (b) $\phi = 5^\circ$ (c) $\phi = 10^\circ$

(d) $\phi = 15^\circ$ (e) $\phi = 20^\circ$ (f) $\phi = 25^\circ$

(g) $\phi = 30^\circ$ (h) $\phi = 35^\circ$ (i) $\phi = 40^\circ$

(j) $\phi = 45^\circ$ (k) $\phi = 50^\circ$ (l) $\phi = 60^\circ$

Figure 7.17: Plots of match values (red) and mismatch values (blue) for different six-view camera configurations. Cameras are positioned in a semicircle with uniform angular spacing. Each configuration has a different elevation angle ϕ, where $\phi = 0$ corresponds to a coplanar camera configuration.

7.6 Summary

It has been demonstrated that attempting to align silhouette set pairs by minimising ET error provides a means for determining whether the pair is a match or a mismatch. Methods for improving the efficiency and the accuracy of the approach have been investigated.

To improve the matching efficiency, equations that make use of an orthographic projection model were derived. The approach is valid for cases in which the stone is small with respect to the distances to the cameras. This is the case with data sets captured using the six-camera setup. The orthographic-based method runs approximately 30 times faster than the perspective-based method.

In an attempt to improve the accuracy of alignment-based matching, measures of inconsistency based on the CIP constraint were investigated. Since the CIP constraint is stronger than the ET constraint, CIP-based methods potentially make use of more information in the silhouettes to discriminate between matches and mismatches. However, unlike the ET error, where pairwise reprojection errors are accumulated, CIP-based methods make use of a cone intersection that is computed from all views simultaneously. This makes the methods more sensitive to noise. Experiments carried out with synthetic data and downsampled real data show the CIP-based errors outperformed ET error in terms of accuracy only for certain operating points of the ROC curve for setups consisting of fewer than five cameras. The CIP-based methods are therefore not considered any further in this thesis.

ET-based alignment was applied to all pairs across runs for the 2-mirror 5-view gravel data set (using a perspective camera model) and the 6-camera garnet data set (using the proposed orthographic approximation). For the gravel data set, all 660 match pairs were found to have substantially lower ET error than any of the 144 540 mismatch pairs. For the garnet data set, all 2460 match pairs were found to have substantially lower ET error than any of the 602 700 mismatch pairs. This indicates that ET-based alignment is an accurate approach for distinguishing between match and mismatch pairs for the types of data and camera configurations considered in this thesis.

Synthetic data sets were used to investigate the effect of different camera configurations and different image resolutions on match and mismatch distributions of ET error. As expected, with insufficient image resolution and too few cameras, there is overlap between the match and mismatch distributions, and such a setup will produce classification errors. Configurations in which optical axes are coplanar, or close to coplanar are observed to result in distribution overlap that does not occur for well-distributed cameras at the same image resolution.

Chapter 8

Dissimilarity from 3D Shape Approximations

8.1 Introduction

Alignment-based matching (as described in the previous chapter) is accurate, yet slow, since time-consuming nonlinear optimisation must be applied for each comparison of two silhouette sets. Although alignment-based matching is not prohibitively slow for verification tasks (i.e., determining whether a single pair of silhouette sets is a match or a mismatch), the method is too slow to apply to all pairings for batch matching. There are n^2 pairings that can be made between silhouette sets from two runs of n stones. A naive approach, where alignment-based matching is applied to all pairings for $n = 1000$ stones, would take almost a week to process assuming approximately 0.5 seconds per pair (100 starting points and an orthographic approximation).

To address this issue, a fast signature-based method for estimating the dissimilarity between two 3D shapes is proposed. The method uses the idea of shape distributions introduced by Osada et al. [103, 104], along with the compact representations of distributions that Rubner et al. [112, 113] refer to as a *signatures*. Likely matches can be identified with the signature-based method so that the more time-consuming alignment-based matching need only be applied to a small number of cases. The signature-based method requires less than one microsecond to assign a dissimilarity value to a silhouette set pair (after once-off preprocessing has computed a signature for each silhouette set). This allows all pairings between two runs of $n = 1000$ stones to be considered in less than one second.

In this chapter, the performance of the proposed signature-based method is quantified in terms of accuracy when applied in isolation (i.e., without alignment-based matching) to verification and identification tasks. The next chapter shows how the signature-based method can be combined with alignment-based matching to efficiently solve the batch matching problem.

Broadly, the proposed signature-based method for computing dissimilarity is carried out as follows. The VEMH is computed from each silhouette set as an estimate of the 3D shape of the convex hull of the corresponding stone. Caliper diameters are sampled from each VEMH in different directions to create a caliper diameter distribution for each silhouette set. The caliper diameter distribution is approximated with a signature consisting of a vector of a small number of elements. Dissimilarity between pairs of signatures is rapidly computed using the earth mover's distance (EMD).

8.2 Related Work

There is a vast body of literature describing different approaches for defining dissimilarity between 3D shapes. Several survey papers compare the different methods [21, 124]. The methods are broadly classified into graph-based and feature-based methods.

Graph-based methods (such as determining the skeleton of an object) are appropriate for complex shapes, and are typically computationally inefficient. Since stones are simple shapes, graph-based methods are not an appropriate means for matching.

Local feature-based methods, such as shape contexts, have been shown to be effective for shape retrieval, even in cases where only a portion of the object is available (partial matching). However, since these methods are typically inefficient, and since local features cannot be accurately estimated from sparse silhouette sets, they were not considered.

Global feature-based methods compare features (such as volume and moments) or distributions of features computed from the 3D shape. Since global features or feature distributions can be rapidly compared, these approaches have been identified by the survey papers as the most efficient approach to matching, and are appropriate for use as pre-classifiers. The shape distribution framework of Osada et al. has been selected as the basis of the method described in this chapter because of its simplicity, efficiency, and success in a range of applications [22, 64, 104]. Although not considered by the original authors, the framework also facilitates the use of the compact signature representation for which dissimilarity between distributions can be rapidly computed.

Because of its speed, the shape distribution framework has been chosen by researchers for specific shape-lookup applications. Canzar and Remy [22] use shape distributions as a faster alternative to alignment-based techniques to look up protein models from a database that are similar to a query model. Comparisons of 353 766 700 protein shapes were completed in less than an hour, with 97% nearest neighbour agreement on class label. Ip et al. [64] use shape distributions to create a query-by-example interface to a CAD database of mechanical parts.

In their original work on dissimilarity measurement using shape distributions, Osada et al. introduce five functions that are used to form shape distributions:

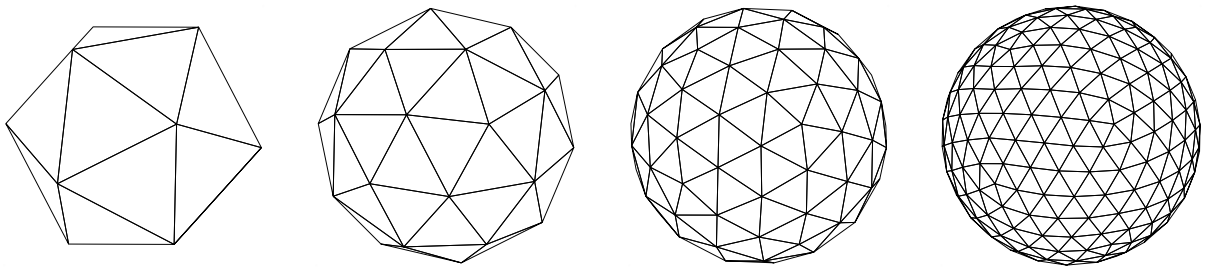
- $D1$ is the distance between a random point on the surface and the centroid;
- $D2$ is the distance between two random surface points;
- $D3$ is the square root of the area of the triangle formed by three random surface points;
- $D4$ is the cube root of the volume of the tetrahedron formed by four random surface points; and
- $A3$ is the angle formed by three random surface points.

The $D2$ shape function was found to be the most accurate shape function for looking up 3D models of everyday objects (cars, humans, phones, mugs), and was found to perform better than feature-based lookup based on moments. In this chapter, these five functions are compared with caliper diameter distributions in terms of matching accuracy.

8.3 Method

The signature-based method uses the VEMH as a 3D approximation of the convex hull of the corresponding stone for each silhouette set. The distribution of caliper diameters over all directions is approximated by sampling caliper diameters in a finite number of directions. Approximately uniform sampling is obtained by using the vertices of a subdivided icosahedron [61] to specify the directions along which to compute caliper diameters.

A subdivided icosahedron of Level L is formed from a subdivided icosahedron of Level $L-1$ as follows. Each face of the Level $L-1$ polyhedron is replaced with a vertex at its centre; all vertices are projected onto the unit sphere, and the resultant convex hull is the Level L subdivided icosahedron. An icosahedron whose vertices lie on the unit sphere is the Level 0 polyhedron. Different levels of subdivision of an icosahedron are illustrated in Figure 8.1.



(a) Level 0 (b) Level 1 (c) Level 2 (d) Level 3

Figure 8.1: Different subdivision levels of an icosahedron.

The same caliper diameters are obtained along directions specified by antipodal vertex pairs of the subdivided icosahedron, so only one vertex per antipodal pair is used.

The dot product of a VEMH vertex $\mathbf{v}$ and the unit direction vector $\hat{\mathbf{d}}$ is used to determine the extent e of the vertex in the direction:

$$e = \mathbf{v} \cdot \hat{\mathbf{d}}. \tag{8.1}$$

The caliper diameter c in direction $\hat{\mathbf{d}}$ is the difference between the maximum and minimum extent values in the direction:

$$c = e_{\max} - e_{\min}. \tag{8.2}$$

Rubner et al. [112, 113] point out that distributions can be efficiently approximated as *signatures* instead of histograms with bins of equal width. Creating signatures involves clustering sample points and representing each cluster with a single point, typically the centroid of the cluster. For one-dimensional distributions, clustering can easily be achieved by selecting histogram bins with equal counts rather than equal widths. The mean values of the sample points in each bin form the signature. By varying the bin width so that bin counts are equal, greater weight is given to describing parts of the shape distribution that have greater density. The signature is therefore a more efficient approximation of the distribution than a histogram.

Figure 8.2 illustrates the process of forming a 10-element signature from a silhouette set for three examples. Note that the silhouette sets in the first two columns correspond to the same stone (however, the stone is oriented differently). The VEMHs and the signatures in the first two columns are therefore similar to one another, whereas the VEMH and the signature from the third column appear dissimilar since they are formed from a different stone.

Dissimilarity between pairs of signatures is computed using the EMD, which is the area between the CDFs (cumulative distribution functions) of the two distributions.

The EMD between two signatures is efficiently computed by directly computing the area between the two CDFs implied by the signatures. The PDFs (probability distribution functions) of the two distributions are approximated by unit Dirac delta functions positioned at the signature element values. The area difference is computed in $O(n)$ time complexity (for n-element signatures) by traversing the two arrays and accumulating the area difference between CDFs. Figure 8.3 illustrates the comparison of signatures between a match pair and a mismatch pair formed from the three examples of Figure 8.2.

Note that Osada et al. [103] investigate various norms between both the PDF and the CDF for measuring dissimilarity between distributions. The compact signature representation is not amenable to computing distances between PDFs, so this approach is not investigated here. The EMD is equivalent to the 1-norm between CDFs. The infinity-norm (i.e., maximum difference) between CDFs is a commonly used dissimilarity measure that is sometimes known as Kolmogorov distance. This is also not investigated here, since the compact signature representations limit the number of discrete values that the Kolmogorov distance could take on. For instance, the dissimilarity between two 2-element signatures could only take on three values: 0, 1 and 2.

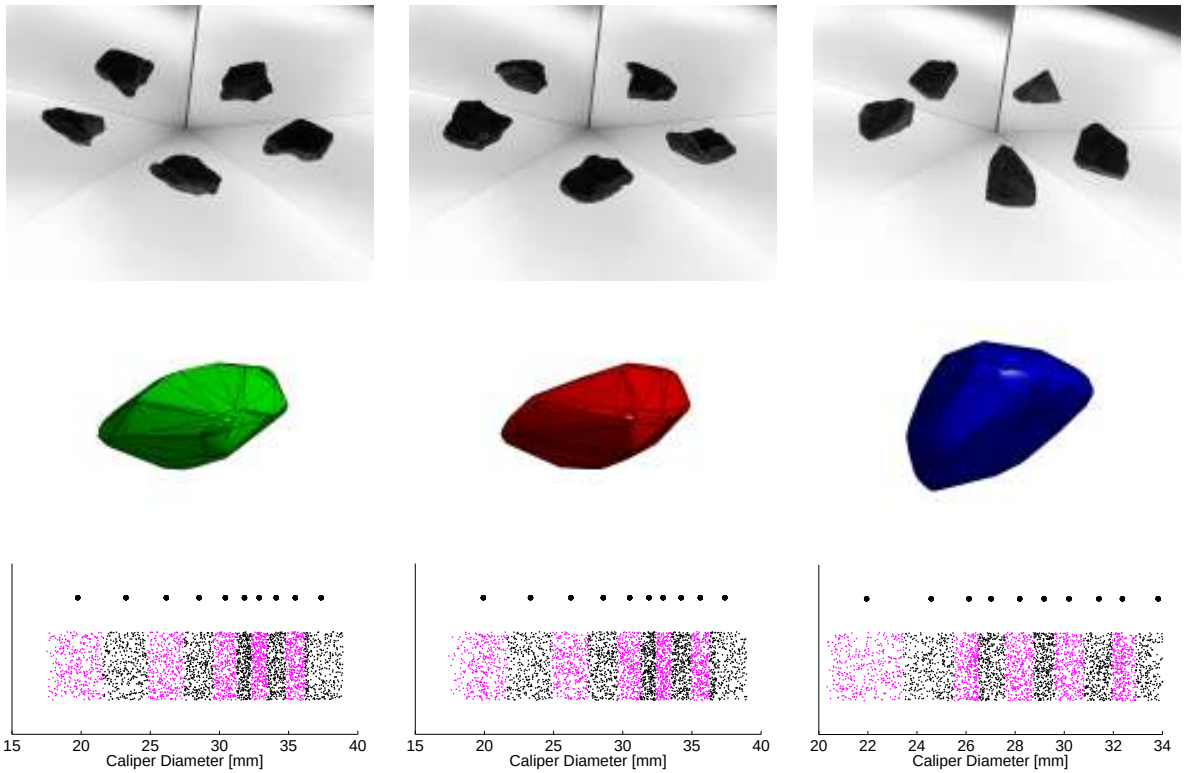


Figure 8.2: Three examples of input silhouette sets *(first row)*, corresponding VEMHs *(second row)*, and caliper diameter sample values and signature values *(third row)*. The vertical component in the caliper diameter value plots is random (a visualisation aid). Larger dots represent signature values computed as the mean of corresponding deciles of caliper diameter values (10 signature elements are used). Decile colouring alternates to show correspondences with signature values. The silhouette set in the first column matches the silhouette set in the second column, but not the silhouette set in the third column. ET-based pose optimisation has been applied to align all VEMHs with the reference frame of the first column.

8.4 Experiments

8.4.1 Numbers of Samples and Signature Elements

Experiments were carried out to investigate the effect of the number of signature elements and the number caliper diameter samples on the matching accuracy achieved using the signature-based method.

Bradley [16] recommends using the area under the ROC curve (AUC) for a single number measure of accuracy. The area under the ROC curve represents the probability that the dissimilarity value associated with a mismatch selected at random will be smaller than the dissimilarity value associated with a match selected at random. Figure 8.4 shows plots of AUC versus number of signature elements for the gravel and garnet data sets. The plots illustrate that further improvements in accuracy are small after approximately ten elements per signature. This indicates that, for the purpose of matching, ten signature elements are able to capture most of the information in the caliper diameter distribution.

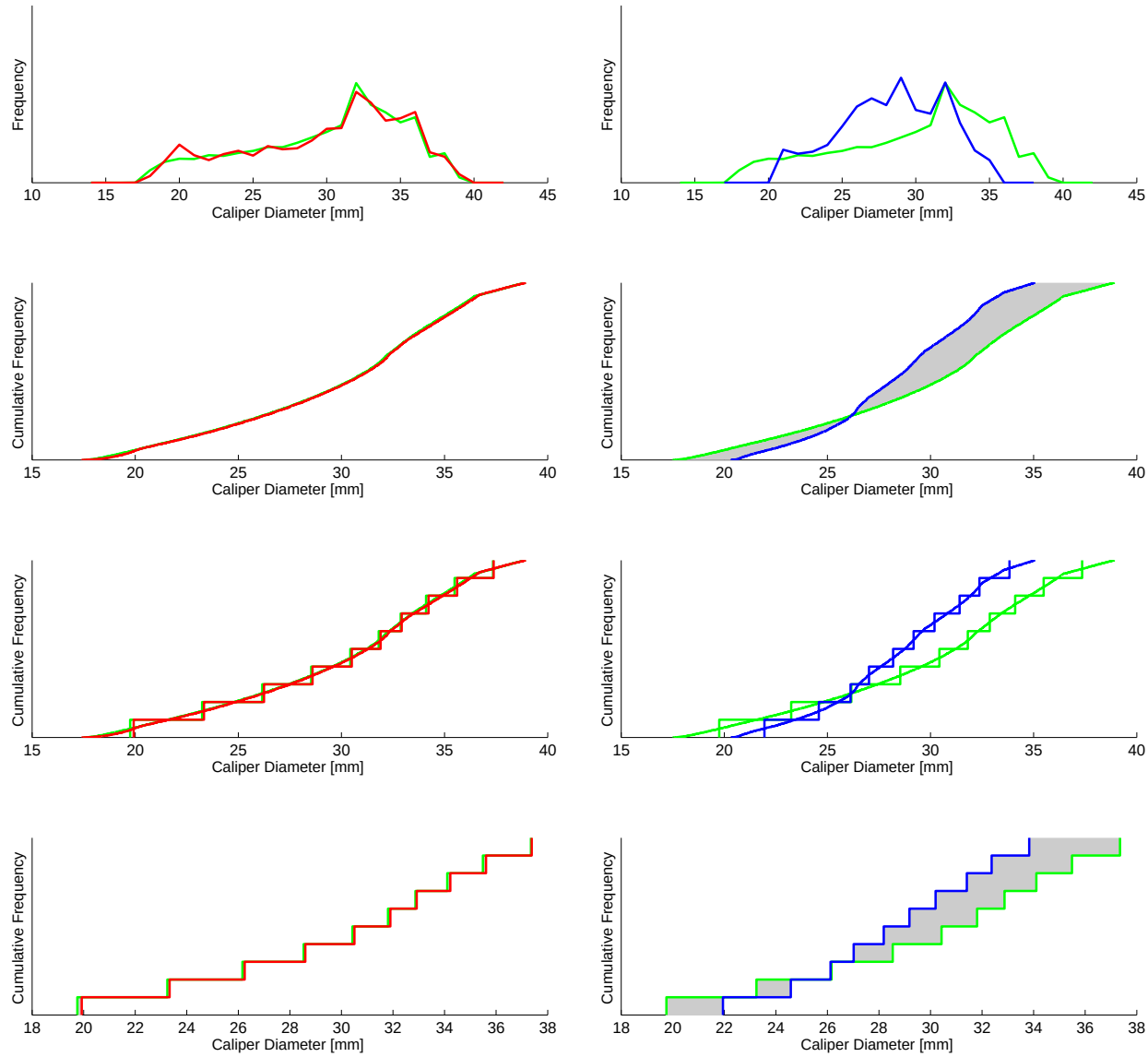
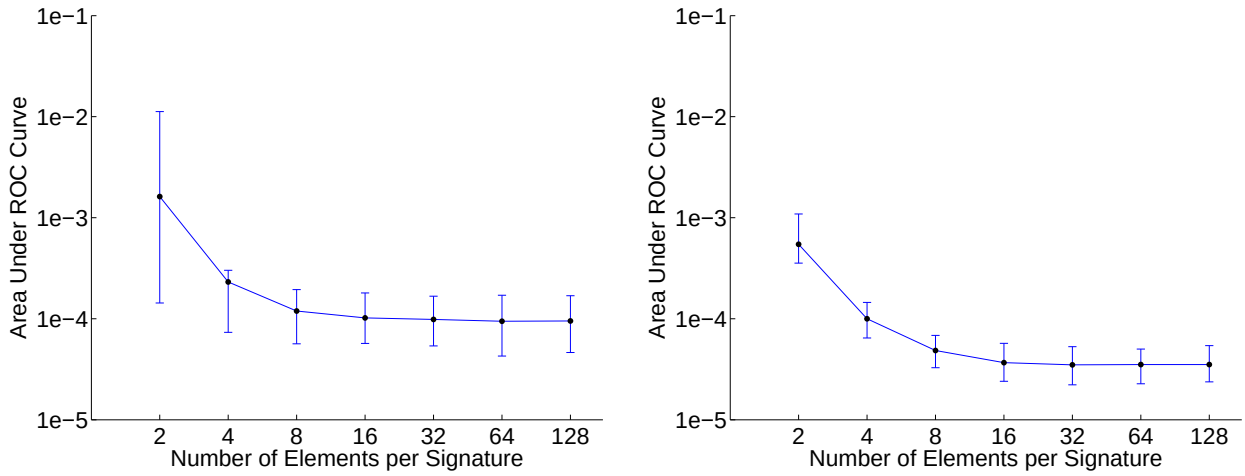


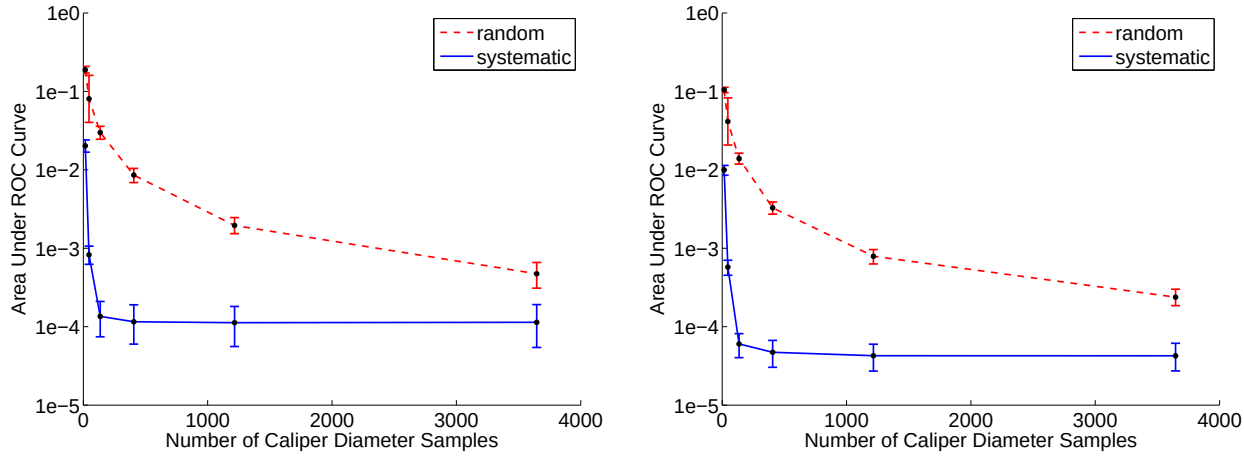
Figure 8.3: Examples of caliper diameter distributions for a match *(first column)* and a mismatch *(second column)*. The match pair is formed from columns one (green) and two (red) of Figure 8.2, and the mismatch pair is formed from columns one (green) and three (blue) of Figure 8.2. The first row shows distributions estimated from the caliper diameter samples using a kernel smoothing method. The second row shows CDFs derived directly from the caliper diameter samples. The area between CDFs (which represents the EMD) is shown in grey. The third row shows the CDFs implied by the signatures overlaid on the original CDFs. The fourth row shows the area between the signature CDFs in grey. This represents the EMD between signatures. The EMD is smaller in the first column (a match pair) than the second column (a mismatch pair).



(a) 2-mirror 5-view gravel data set

(b) 6-camera garnet data set

Figure 8.4: Plot of number of signature elements versus area under the ROC curve for (a) the 2-mirror 5-view gravel data set, and (b) the 6-camera garnet data set. Error bars represent 95% confidence intervals computed using the subsets bootstrap. Six icosahedron subdivisions were used to compute 3646 directions for caliper diameter samples from each VEMH. Note that nonlinear axes have been used.



(a) 2-mirror 5-view gravel data set

(b) 6-camera garnet data set

Figure 8.5: Plot of number of caliper diameter samples versus area under the ROC curve for (a) the 2-mirror 5-view gravel data set, and (b) the 6-camera garnet data set. Error bars represent 95% confidence intervals computed using the subsets bootstrap. Curves have been computed using both systematic sampling using subdivided icosahedra, and random sampling. The following numbers of samples were used: 16 (corresponding to one subdivision of an icosahedron), 46 (two subdivisions), 136 (three subdivisions), 406 (four subdivisions), 1216 (five subdivisions), and 3646 (six subdivisions).

Figure 8.5 shows the results of an experiment in which the number of caliper diameter samples used to estimate each distribution is varied.

Results are shown for both systematic sampling (based on icosahedron subdivision) and random sampling (using a uniform random distribution of points on a sphere). The results clearly indicate that systematic sampling outperforms random sampling using the same number of samples. Little further improvement is

observed with more than 406 samples (from four icosahedron subdivisions) in the case of systematic sampling. Subsequent experiments therefore make use of 10-element signatures formed from 406 systematically selected caliper diameter samples.

Using these parameter values, a 3.2 GHz Pentium 4 machine takes an average of 1.3 milliseconds to compute the caliper signature from the VEMH for silhouette sets captured using the six-camera setup. Computing the EMD between two signatures takes an average of 0.5 microseconds.

8.4.2 Comparison with ET Error

The EMD between signatures was computed for all match and mismatch pairs for the garnet data set and compared with the ET error across the same pairs. A plot of EMD versus ET error is shown in Figure 8.6. The EMD and ET error values are highly correlated with one another. The closeup in Figure 8.6b shows that whereas the ET error separates all match pairs from mismatch pairs, the EMD between signatures does not. The EMDs are however substantially faster to compute than the ET errors: the EMD between signatures take approximately half a microsecond to compute and the ET errors take approximately half a second. The EMDs are therefore faster to compute by a factor of a million.

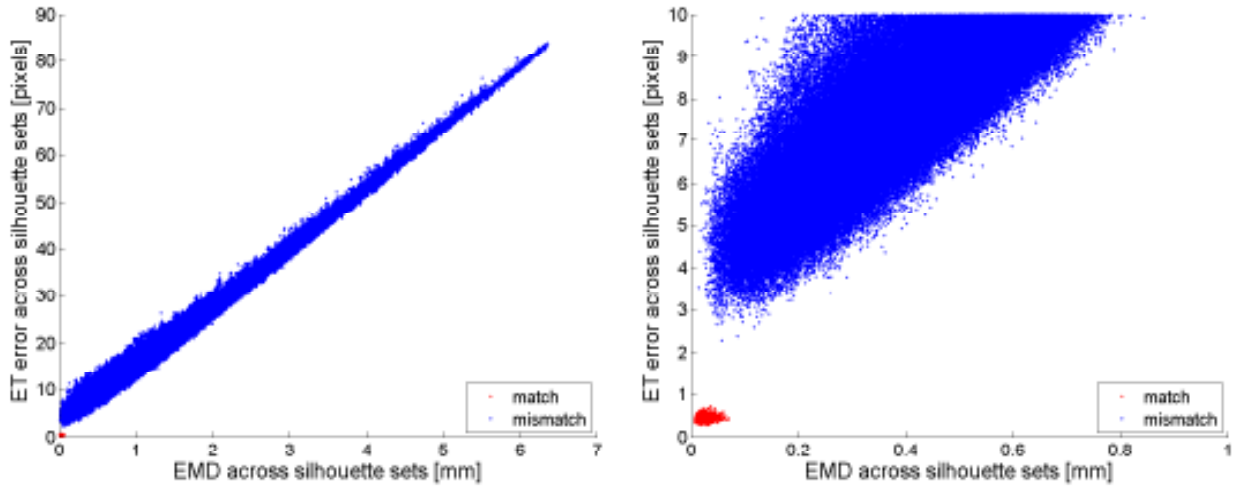
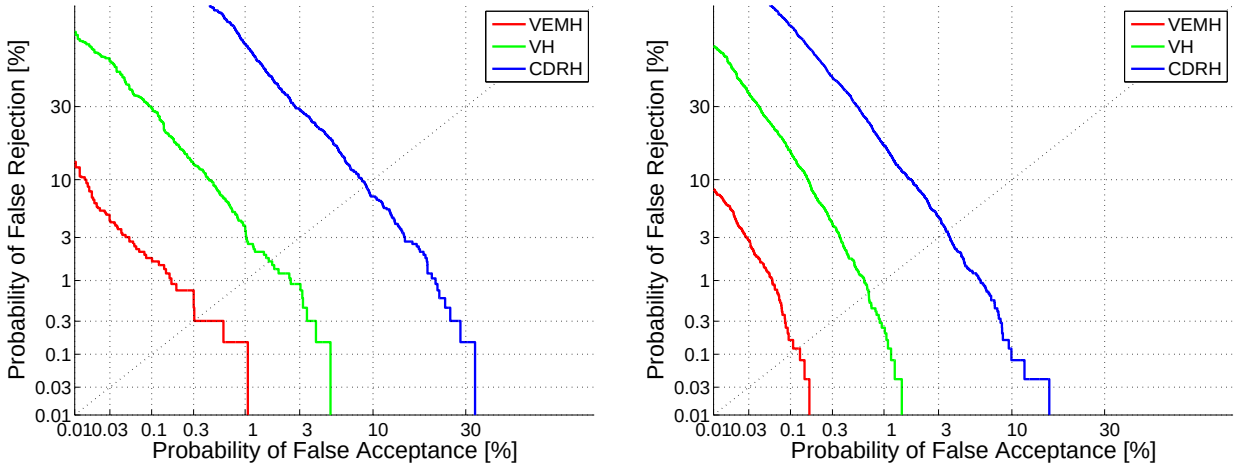


Figure 8.6: Plot of EMD versus ET error for silhouette set pairs from the 6-camera garnet data set. The plot on the right is a closeup of part of the plot on the left.

8.4.3 Different Methods of Estimating Stone Shape

Caliper diameter signatures were computed using the visual hull and the constant depth rim hull (CDRH) as alternatives to the VEMH for estimating the convex hull of the stone from its silhouette set. The ROC curves shown in Figure 8.7 illustrate that greater accuracy is achieved using the VEMH than the two competing methods. The CDRH produces the worst results. Efron's method [35] of visualising the statistical variability



(a) 2-mirror 5-view gravel data set

(b) 6-camera garnet data set

(c) 2-mirror 5-view gravel data set

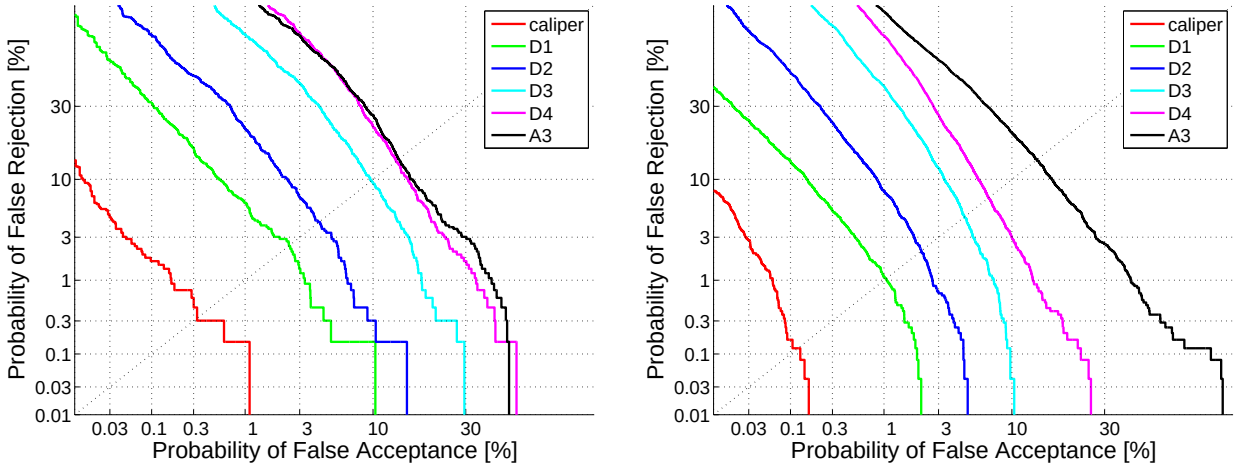
(d) 6-camera garnet data set

Figure 8.7: ROC curves derived from (a) the 2-mirror 5-view gravel data set, and (b) the 6-camera garnet data set for caliper signatures computed using different means to approximate the 3D convex hulls of stones: VEMHs, visual hulls, and CDRHs; (c) twenty bootstrap curves drawn from the data presented in (a); (d) twenty bootstrap curves drawn from the data presented in (b).

associated with the curves is illustrated in Figures 8.7c and 8.7d: the subsets bootstrap [10] is used to show 20 bootstrap replications (an estimate of what would be observed if the experiment were repeated 20 times with new samples). The bootstrap replications indicate that the differences between the curves is sufficiently low that the observed differences cannot be attributed to chance. The greater variability present in the bottom right of the curves is an artefact caused by using many more mismatch pairs than match pairs. Part of the reason that the plots for the gravel data exhibit greater variability than the plots for the garnet data is that fewer runs were used (3 runs that provide 3 pair combinations across runs versus 5 runs that provide 10 pair combinations across runs).

8.4.4 The Shape Functions of Osada et al.

Figure 8.8 shows ROC curves computed using the caliper diameter distribution and the shape functions suggested by Osada et al. [104]. The caliper diameter distributions outperform all of the shape functions of Osada et al. for both data sets.



(a) 2-mirror 5-view gravel data set

(b) 6-camera garnet data set

Figure 8.8: ROC curves derived from (a) the 2-mirror 5-view gravel data set, and (b) the 6-camera garnet data set for caliper signatures and for the shape functions proposed by Osada et al. [104]. One million samples of each shape function of Osada et al. was used for each VEMH. This value was found to be sufficiently large so that further increases showed negligible improvement in accuracy.

The shape functions show a wide range of performance, with the distance-based (DN) features degrading as the number N of random surface points used to compute each sample is increased. The worst performing shape function is the $A3$ feature, which is based on angle distributions rather than distance-based distributions.

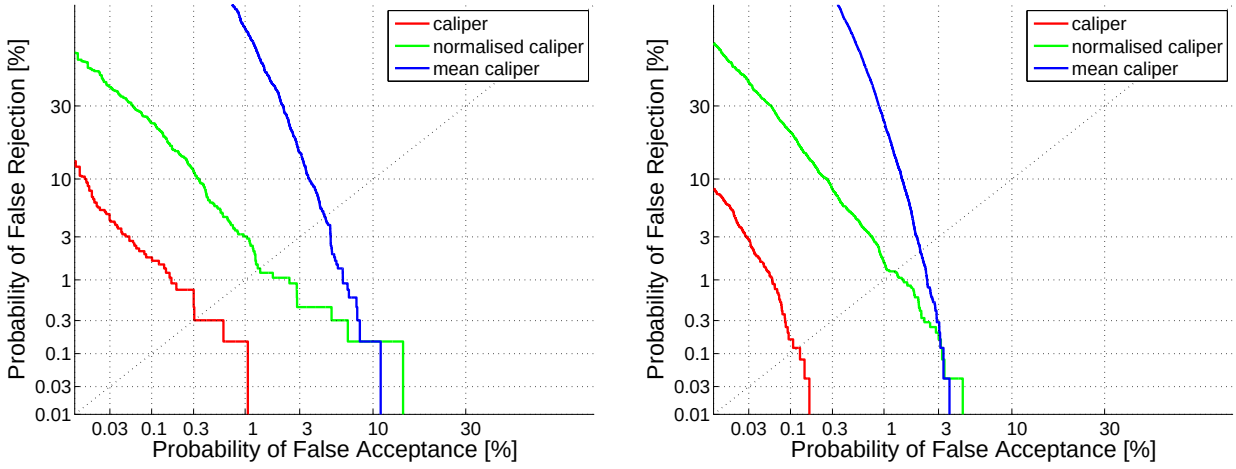
Note that Osada et al. use the functions to identify similar shapes from existing mesh models. Here, dissimilarity is based on approximate 3D shapes that are derived from silhouette sets. The more accurate

performance of the caliper distribution indicates that the caliper distribution of a stone can be more accurately inferred from the silhouette sets than Osada et al.'s shape functions (with respect to the variation of the shape distribution amongst different stones). This does not imply that caliper distributions would outperform Osada et al.'s shape functions for the 3D model retrieval application for which they were designed.

8.4.5 The Effect of Size and Shape

Part of the ability of the signature-based method to distinguish match pairs from mismatch pairs is the size variability of the stones within each data set. To obtain an indication of the performance of the signature-based method with only shape information, matching was carried out using normalised signatures. Normalisation was carried out by dividing each caliper diameter distribution by its mean. In addition, matching was carried out using only size information: the mean diameter values were used as 1-element signatures.

Figure 8.9 shows the results in terms of ROC curves. As expected, the normalised signatures provide lower



(a) 2-mirror 5-view gravel data set

(b) 6-camera garnet data set

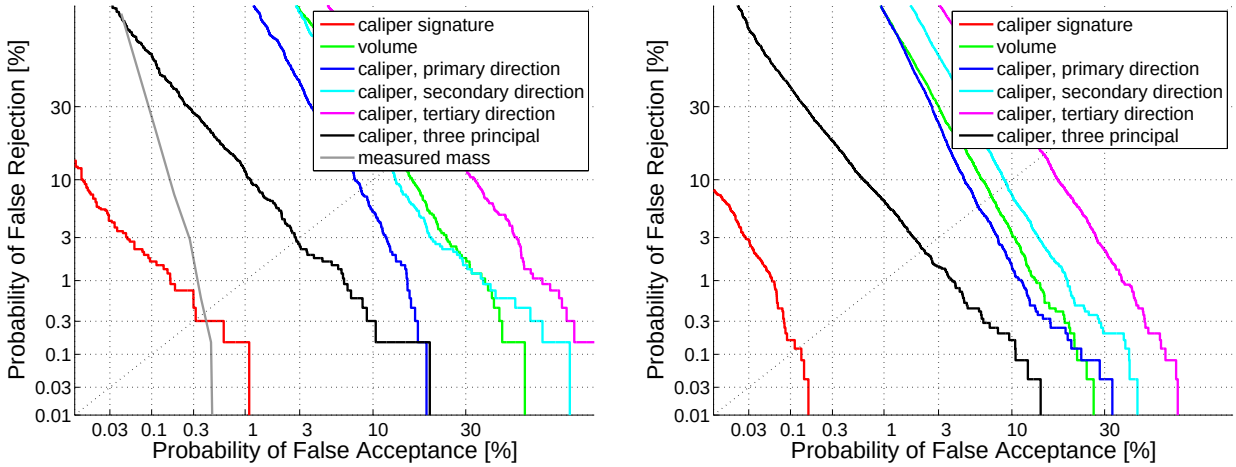
Figure 8.9: ROC curves derived from (a) the 2-mirror 5-view gravel data set, and (b) the 6-camera garnet data set for caliper signatures, normalised caliper signatures, and mean caliper diameter values. Normalised caliper signatures are normalised by dividing by the mean caliper diameter to create signatures with unit mean in all cases. This demonstrates the accuracy obtainable without scale enforcement after camera calibration, or equivalently, the accuracy obtainable with shape information but not size information. The mean caliper diameter shows the accuracy obtainable with size information but not shape information.

accuracy than the original signatures, since size information has been discarded. However, the normalised signatures outperform the mean caliper signatures for most operating points. This indicates that the caliper signatures accurately capture some essence of stone shape, rather than discrimination accuracy being due to the size variability present in the data sets. The plots indicate, for example, that an operating point can be chosen (for either data set) so that the equal error rate is approximately 2%. This means that, for a certain EMD threshold, the signature-based method would correctly classify a randomly selected match or mismatch pair with 98% probability *without* scale information.

8.4.6 Feature-Based Dissimilarity

The signature-based method was compared with a few simple feature-based methods. To justify its additional complexity over the simpler feature-based methods, the signature-based method should provide superior accuracy.

Four features were measured from each VEMH: volume and caliper diameters along the three principal directions. The absolute difference between the two feature values associated with each pair was used as a measure of dissimilarity. In addition, the Euclidean distance between a 3-vector consisting of all three principal caliper diameters was used as a further simple feature-based method. ROC curves based on the different methods are shown in Figure 8.10. The plots indicate that the signature-based method substantially



(a) 2-mirror 5-view gravel data set

(b) 6-camera garnet data set

Figure 8.10: ROC curves derived from (a) the 2-mirror 5-view gravel data set and (b) the 6-camera garnet data set for caliper signatures and feature-based measurements. Dissimilarity is defined by difference in volume, and difference in caliper diameter measured along the three principal directions of the VEMH (primary, secondary, and tertiary). Dissimilarity defined as the Euclidean distance between a 3-vector of the three caliper diameters along the principal directions (three principal) is shown. The ROC curve computed using differences between mass measured on an electronic balance is shown for the gravel data set.

outperforms the simple feature-based methods. Caliper diameter along the tertiary principal direction (short diameter) is the worst performing feature. Caliper diameter along the primary principal direction (long diameter) outperforms caliper diameters measured along the other two principal directions. This is consistent with observations made in Chapter 3 that indicate that a large degree of variability is associated with estimating short diameters from silhouette sets. Using all three diameters provides better performance than any one diameter.

The plot in Figure 8.10a also shows the ROC curve derived from the gravel masses measured using an electronic balance. The plot does not appear stepped like the other curves. This is because of the discretised nature of the mass measurements: more than one mass measurement difference corresponds to the same value. The ROC curve of the mass measurements crosses the ROC curve of the signature-based method,

indicating that which method is better depends on the operating point selected. The mass-based approach performs poorly at low probability of false acceptance. This is because there are 61 mismatch pairs (of the 144 540 mismatch pairs considered) whose dissimilarity values (that is, differences between the measured masses of two different stones) are exactly zero. (The resolution of the electronic balance was 0.01 grams; the mean mass was 20.72 grams, and the standard deviation of the mass values was 6.91 grams.)

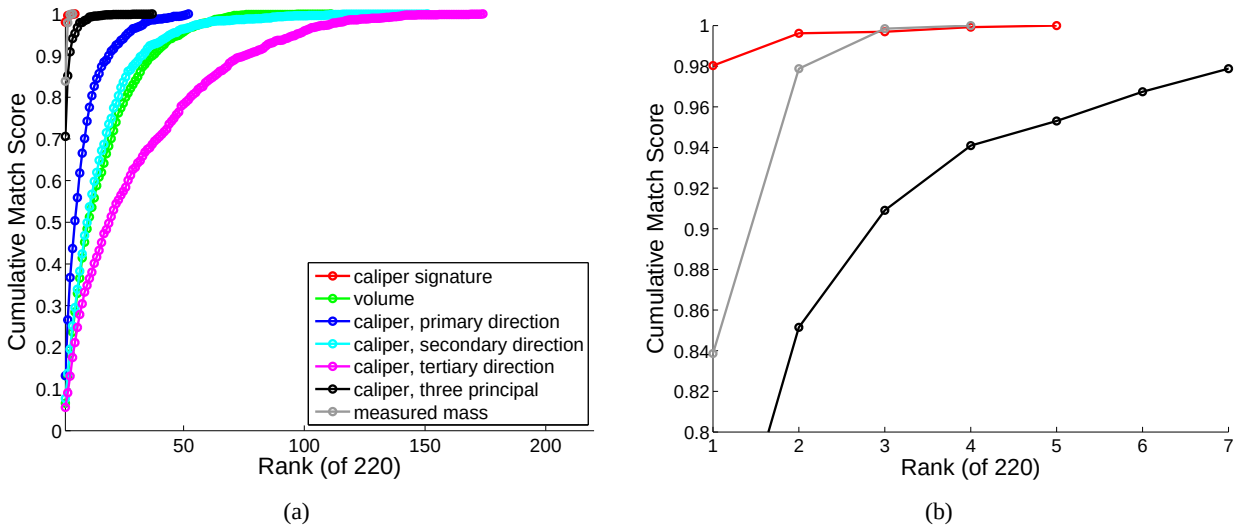


Figure 8.11: (a) Rank versus cumulative match score plot derived using various measures of dissimilarity; (b) a closeup of part of the plot shown in (a).

Rank versus cumulative match score plots [106] show the proportion of cases in which a query ranks within the top r matches. For instance, a rank of $r = 5$ with a cumulative match score of 0.85 means that the correct match ranks amongst the top five matches (ordered from smallest to largest EMD) in 85% of all cases. The plots are computed by considering each case as a query in turn, and comparing each query with the other cases from another run. All combinations of runs are considered, with cases from each run being considered as queries and as database entries. Rank versus cumulative match score plots are useful for quantifying performance in closed universe [106] scenarios, where the query is known to match one of a certain number of database entries.

Figure 8.11 shows the rank versus cumulative match score plot derived from the gravel data. This provides an indication of how well the signature-based approach performs at the task of *identifying* a stone from a database of 220 pre-stored silhouettes sets, one of which is known to match the query silhouette set. A practical system could use alignment-based matching to classify database-query pairs in an order specified by signature-based dissimilarity. The plot indicates that the probability of the first pair considered by alignment-based matching being a match is 98%.

The plot shows that the signature-based approach outperforms the feature-based approaches. Although the signature-based method is more likely than the measured mass method to contain the match in the pairs

ranked up to one and two, the measured mass method is more likely to contain the match in the pairs ranked up to three and four.

8.5 Summary

This chapter has proposed a simple method based on caliper distribution signatures for computing a measure of dissimilarity between silhouette sets. The signatures and the dissimilarity between signatures can be rapidly computed.

The method achieves its efficiency by using approximations to 3D shape, rather than relying on silhouette consistency constraints. This approach places an inherent limitation on the accuracy that can be achieved using the method, since there are inherent ambiguities in inferring 3D shape from a sparse silhouette set.

Caliper distribution signatures have been shown to outperform simple feature-based methods (such as volume) as well as the five approaches introduced by Osada et al.

Since the method facilitates rapid ranking of silhouette sets in order of similarity, the signature-based method can be used in conjunction with the alignment-based method described in the previous chapter to identify a stone from a query silhouette set, by matching a previously stored silhouette set in a database.

Chapter 9

Batch Matching

9.1 Introduction

Batch matching is finding the one-to-one correspondences between silhouette sets from two unordered runs of the same batch of stones: each silhouette set in the first run must be matched to the silhouette set in the second run that was produced by the same stone.

This is a *square assignment problem* since each of the n objects in the first run must be matched to one of the n objects in the second run. The matching can be specified by an $n \times n$ *permutation matrix* in which each element is either one or zero (indicating match or mismatch), and each row and each column sums to one.

The proposed approach to batch matching makes use of the desirable characteristics of the two measures of dissimilarity developed in Chapters 7 and 8: alignment-based matching, where ET error is the measure of dissimilarity, and signature-based matching, where the EMD between signatures is the measure of dissimilarity. The desirable characteristic of alignment-based matching is its accuracy, whereas the desirable characteristic of signature-based matching is its speed.

Signature-based matching is used to compute a measure of dissimilarity between all pairings of silhouette sets in the first run with those in the second run: for n stones there are n^2 pairings. Prior knowledge of the distributions of dissimilarity values for match and mismatch pairings is used to estimate likelihood ratios for each pairing (indicating the likelihood of being a match). Pose optimisation is then successively applied to the pairing with the greatest likelihood ratio. If pose optimisation from a given starting point (initial pose estimate) leads to a sufficiently low error, then the pairing is labelled a match and is removed from consideration. Otherwise the likelihood ratio is *updated* to reflect that a failed pose optimisation from the given starting point indicates that the pairing is less likely to be a match. Starting points based on the principal axes of 3D approximations to the stone are used, followed by uniform random orientations. The proposed greedy algorithm (which processes the pairing with the greatest likelihood ratio at each iteration)

is demonstrated to produce close to optimal performance on a test set of six-view silhouettes of 1200 uncut gemstones (i.e., the time spent processing mismatches is small). On a 3.2 GHz Pentium 4 machine, the once-off per silhouette set computations take approximately 50 seconds. Computing dissimilarity between shape distributions takes 0.7 seconds and pose optimisation takes 17 seconds (of which 15 seconds is spent considering matches and 2 seconds on mismatches).

There are n^2 comparisons (or pairings) that can be made between silhouette sets in the first run and silhouette sets in the second run. Although the proposed algorithm is still inherently of at least $O(n^2)$ time complexity, the n^2 component dominates only for very large n. This means that batches of more than a thousand stones can be matched efficiently.

The batch matching algorithm makes use of several key ideas that together achieve efficiency:

1. **Shape distribution dissimilarity for ranking pairings by likelihood of match.** EMDs are computed between estimated caliper diameter distributions for each of the n^2 pairings between first run silhouette sets and second run silhouette sets. (EMDs are computed efficiently, taking less than a microsecond per pairing.) Likelihood ratios are computed for each pairing from the EMD using prior knowledge of distributions of EMDs for match and mismatches. A priority queue is used to access pairings so that the most likely matches can be processed first.

2. **Recomputing the most likely match after pose optimisation from one starting point.** Pose optimisation proceeds by optimising from a single pose estimate at a time. After pose optimisation, the likelihood ratio is updated if the associated ET error is above the threshold for matches. (Knowing that a pose optimisation fails from a given starting point implies that a match is less likely than before this is known). The pairing is pushed back into the priority queue with its updated likelihood ratio. If the likelihood ratio has been decreased by a sufficiently small amount, then the pairing will remain at the front of the priority queue, otherwise a new pairing will be selected for processing. This approach ensures that ET-based pose optimisation is always applied to the pairing that is most likely a match (based on EMD between signatures and number of failed pose optimisations so far).

3. **Certainty of a match implies certainty of mismatches.** If ET-based pose optimisation leads to an ET error that is below the match threshold, the pairing is labelled as a match (i.e., the probabilistic framework is abandoned and a hard decision is made). This means that all other pairings associated with the two matched silhouette sets can be labelled as mismatches and removed from consideration. This amounts to zeroing the remaining permutation matrix elements that share a row or a column with the matched element. In other words, finding a match implies that mismatches have been found too. (Although possibly obvious, this removal of mismatches from consideration is an important factor in substantially reducing the running time of problems in which a one-to-one correspondence exists, and is therefore explicitly mentioned.)

4. **Good starting points provided by moments of 3D shape approximations.** Pose starting points are selected using the principal axes of a 3D approximation to the stone as a guide. The first pose starting point aligns the principal axes of the two 3D approximations and ensures that their third order moments have the same sign. This starting point leads to an ET error below the match threshold in approximately 80% of match cases. The next three pose starting points align the principal axes in the three other possible ways. Subsequent pose estimates align the centroids of the 3D approximations and select the orientation component using a uniform random rotation.

9.2 Approach

This section describes the greedy algorithm that was designed to efficiently solve the one-to-one correspondence problem for silhouette sets.

9.2.1 Design Rationale

The proposed algorithm is based on the assumption that all matching pairs can be aligned so that the ET error across the two silhouette sets is below a fixed threshold value, and that no mismatch pairs can be aligned so that ET error is below the threshold. This assumption is valid if noise levels are sufficiently low, and stone shapes are sufficiently dissimilar. (Section 9.3.2 demonstrates the consequences of using a data set for which the assumptions do not hold.) The threshold must be determined from a training data set.

The aim of the algorithm is to find the n silhouette set pairs with ET errors below the threshold. Once pose optimisation has determined pose parameters that align n pairs sufficiently well (i.e., ET error across the silhouette set pair that is below the threshold), the algorithm terminates, since the one-to-one correspondence has been determined.

ET-based pose optimisation is time-consuming, and for an efficient matching algorithm it must be kept to a minimum. Efficiency is achieved by combining two strategies:

1. As little time as possible is spent on pose optimisation between pairs that do not match.
2. As little time as possible is spent optimising from starting points (initial pose estimates) that lead to insufficiently low ET errors (local minima) for pairs that do match.

The first strategy is implemented by selecting, in each iteration, the pairing most likely to match (based on the information considered: EMD and number of optimisations failed so far). Since the most likely match is selected at each iteration, rather than trying to minimise total running time, the proposed algorithm is *greedy*.

The second strategy is carried out by using the principal axes and moments of 3D approximations to each stone to select the starting points that are most likely to lead to the correct alignment of matching pairs.

Figure 9.1 shows a flow chart of the greedy algorithm. Note that the algorithm could finish after $n-1$

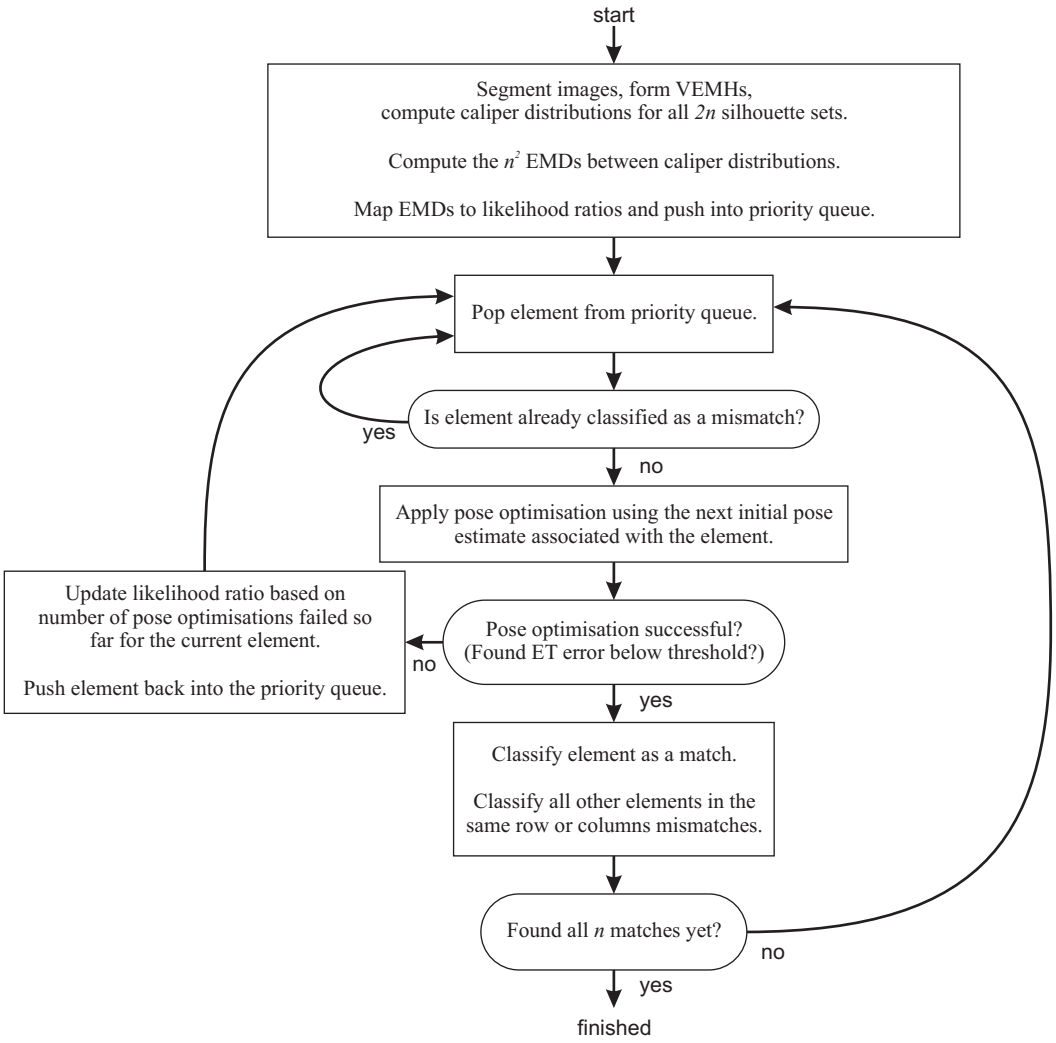


Figure 9.1: Flow chart for the proposed greedy algorithm.

matches are found, since the single remaining unmatched element must be the match. Instead, all n elements are matched with ET-based pose optimisation. This results in a very small increase in total running time.

9.2.2 Initial Likelihoods from EMDs

EMDs between caliper diameter signatures are computed for each of the n^2 pairings between silhouette sets in the first run and silhouette sets in the second run.

The EMDs are used to *order* the pairings so that pairings that are the most likely to match are considered first. A reasonable approach would be to process the pairings (apply pose optimisation) in the order specified by the raw EMD values. This would, however, require the number of pose optimisations to be considered for each pairing to be specified in advance. Alternatively, a single pose optimisation could be applied for each pairing in turn up to a certain threshold on the EMD, after which pairings are reconsidered in turn from different starting points.

A better approach, however, is to update the likelihood of a pair being a match based on the additional information of the number of failed optimisations that have been carried out for the pair. (A failed optimisation is one in which a pose with an associated ET error below the match threshold could not be found.) By using additional information, matches are more likely to be selected than mismatches than if only the raw EMD values were used. This requires that the EMD values be mapped to likelihood ratios so that they can be updated using Bayes's rule.

9.2.3 Training

To determine the mapping from EMD values to likelihood ratios, a training set is required. A training set consists of multiple runs of silhouette sets of a batch of stones for which the correspondence between silhouette sets is known. The training set is a random sample from the population of stones for which the batch matching is to be used. The ratio of match density to mismatch density must be estimated for all EMD values. Many methods exist for estimating probability density from samples [34]. This problem also has the additional constraint of monotonicity: a greater EMD implies a lower likelihood of match. Arandjelović describes a method to enforce the constraint of monotonicity [2]. A simple histogram method, however, was found to produce good results, so more sophisticated methods were not implemented. A coarse histogram (five bins) was formed for EMD values from match pairs in the training data and for mismatch pairs in the training data. Ratios of normalised bin counts at the bin centres were used to form a mapping from EMD values to likelihood ratios. Piecewise linear interpolation was used to determine values between the bin centres.

The training procedure also uses the training set to determine a threshold value on the ET error for matches, and to estimate the probability of failed alignment for a match pair after s starting points have been used.

To determine a threshold value, the largest ET error across silhouette sets for a match, and the smallest ET error across silhouette sets for a mismatch are estimated. The threshold is chosen to be midway between these two values.

Since applying pose optimisation from many starting points to all training set pairs is too time-consuming, the following method was used. The mismatch pairs are ordered by EMD between signatures, and pose optimisation is only applied to the first 1000 mismatch pairs. Pose optimisation is applied to the cases only from the four starting points specified by principal axis alignment. This approach ensures that an ET error

value that is approximately as small as the smallest mismatch error can be computed in a reasonable running time.

The estimate of the largest ET error for match pairs is computed by applying pose optimisation from the four principal axis starting points to each match pair. Pose optimisation is then applied from a random starting point to the match pair whose lowest ET error is the largest. This is repeated until the match pair with largest minimum ET error, has had pose optimisation applied from 1000 random starting points. This match pair is used to estimate the largest ET error across silhouette sets for a match.

Once the threshold has been specified, the proportion of match cases that lead to an ET error below the threshold value is computed for starting points based on the principal axes, followed by random starting points.

9.2.4 Forming a Priority Queue

Pairings are stored in a priority queue that is prioritised by the value specifying the likelihood of match. The indices of the silhouette sets that make up the pair are also associated with each element in the priority queue. These indices are used to reference a permutation matrix that is built up as the algorithm progresses. When a match is found, the corresponding permutation matrix element is changed from ‘unknown’ to one, and other elements in the same row and column are zeroed. When an element that references a zero in the permutation matrix is at the front of the priority queue, it is popped from the queue and no pose optimisation is applied since the pair is already known to be a mismatch. The number of failed optimisations that have been applied to the pair is also associated with each element in the priority queue.

9.2.5 Pose Optimisation

Pose optimisation is applied to the pair of silhouette sets associated with the front of the priority queue (provided that this element has not already been labelled as a mismatch, in which case it is popped and the next element is considered). Pose optimisation attempts to determine the relative pose between two silhouette sets with the assumption that the sets were produced by the same stone.

The starting points (initial pose estimates) for pose optimisation are based on the principal axes and moments of inertia of the VEMHs from each silhouette set. Choosing the pose that aligns the principal axes of the 3D stone approximations and that ensures that the third order moments have the same signs, leads to an ET error below the match threshold value in approximately 80% of cases when the pair is a match. A pose in which the translational component of pose is chosen so that centroids from the two 3D approximations coincide, and the rotational component is a uniform random rotation, leads to the correct alignment in only about 10% of cases. After considering all four possible pose alignments based on the principal axes, the correct alignment is found in all but about 2% of cases (as illustrated in Figure 6.8 on page 111).

The first four initial pose estimates for a pair are therefore chosen to correspond to the four poses that align the principal axes. The pose that keeps the signs of the third order moments unchanged is first. The second two poses change only one sign of the third order moments. After four poses have been considered, uniform random orientations are used for the following poses. A systematic orientation sampler described in the robotics literature [142] was considered, but some initial experimentation showed no evidence of better results.

Note that the one-to-one matching constraints remove the need to decide on the number of pose optimisations to apply: optimisations are applied until all the matches are found. Compare this with the situation of searching for a tag stone that may or may not be present in a batch of stones: in this case a decision must be made to stop applying pose optimisation after it has been applied from a certain number of starting points.

9.2.6 Updating Likelihood Values

After a failed optimisation, the likelihood ratio associated with a pair is updated to reflect both the associated EMD value and the number of failed optimisations.

The probability of a match given a certain number of starting points from which optimisation has been applied must be estimated from a training data set. It is assumed that optimisation will always fail with mismatch pairs. The proportion of cases that fail after one, two, three, and four pose optimisations is computed from the training data (using the pose ordering as described in Section 9.2.5). If the probability of failed pose optimisation from a single pose with a random orientation component is p_1, the probability p_m of failure for all of m random starting points is

$$p_m = p_1^m. \tag{9.1}$$

The value of p_1 will vary for different silhouette set pairs. As an approximation, the mean value of p_1 is estimated from the training set, and Equation 9.1 is used to estimate p_m.

The posterior odds $P(H_{\text{match}}|\text{data})/P(H_{\text{mismatch}}|\text{data})$ of an element being a match is given by Bayes's rule:

$$\frac{P(H_{\text{match}}|\text{data})}{P(H_{\text{mismatch}}|\text{data})} = \left(\frac{P(\text{data}|H_{\text{match}})}{P(\text{data}|H_{\text{mismatch}})}\right)\left(\frac{P(H_{\text{match}})}{P(H_{\text{mismatch}})}\right), \tag{9.2}$$

where H_{match} is the match hypothesis and H_{mismatch} is the mismatch hypothesis. Note that the prior odds $P(H_{\text{match}})/P(H_{\text{mismatch}})$ are the same for all elements (the stones are assumed to be in random order), so ordering by the likelihood ratio $P(\text{data}|H_{\text{match}})/P(\text{data}|H_{\text{mismatch}})$ is the same as ordering by the posterior odds.

The updated likelihood ratio r_{updated} is computed from new observations as follows:

$$r_{\text{updated}} = \frac{P(\text{data}|H_{\text{match}})}{P(\text{data}|H_{\text{mismatch}})} r_d, \tag{9.3}$$

where r_d is the likelihood ratio computed from the EMD value. Here, the data specifies the number of failed optimisations so far. Since $P(\text{data}|H_{\text{mismatch}}) = 100\%$,

$$r_{\text{updated}} = p_s r_d, \tag{9.4}$$

where p_s is the proportion of match cases for which optimisation fails in all cases after using s starting points. For $s > 4$, p_s is estimated using

$$p_s(s) = p_s(4)p_1^{s-4}. \tag{9.5}$$

Note that likelihood ratios are computed without using the one-to-one correspondence constraint. Making use of this constraint does not aid efficiency, since each pairing requires evaluation of a function of values associated with all other pairings. Knowledge of the one-to-one constraint is therefore discarded, and likelihood ratios are computed without considering values associated with other pairings.

9.3 Experiments

This section describes a set of experiments that were carried out using a C++ implementation of the proposed algorithm. The experiments aim to quantify the behaviour of the proposed algorithm in terms of running time, and to quantify the relative importance of the various components in keeping the running time as small as possible.

Experiments were carried out using a data set of 1423 uncut gemstones (pictured in Appendix C, pages 222–224). Ten runs of six-view image sets were captured, yielding a total of $1423 \times 10 \times 6 = 85\,380$ images. Computations were carried out on a 3.2 GHz Pentium 4 machine. For each trial, runs corresponding to 243 randomly selected stones were used as a training set, leaving the remaining 1200 stones as a test set. All 45 run pair combinations of 243 stones were used for training, providing 10 935 match pairs and 2 646 270 mismatch pairs across runs. For each trial, two runs were selected at random from the ten available runs to form a test set of two runs of 1200 silhouette sets.

9.3.1 Preprocessing Running Time

Table 9.1 gives a breakdown of the mean running time for the various preprocessing components. Ten signature elements were computed for each silhouette set using four subdivisions of an icosahedron to determine the caliper sampling directions. The results show a mean processing time of 20.7 ms per silhouette set. The once-off preprocessing per silhouette set is therefore sufficiently fast that it can be carried out online as the stones are fed through the six-camera setup at a rate of ten stones per second.

Computation	Running Time	Percentage of Total
segmentation	9.0 ms	43.5%
convex viewing edges	5.3 ms	25.6%
3D convex hull	3.2 ms	15.5%
2D convex hulls	1.5 ms	7.2%
caliper signatures	1.3 ms	6.2%
moments	0.3 ms	1.4%
edge angle data structure	0.1 ms	0.5%
Total	20.7 ms	100%

Table 9.1: Mean running time for preprocessing a 6-view silhouette set.

9.3.2 Batch Matching with the Proposed Greedy Algorithm

The proposed batch matching correctly matches silhouette sets across two runs of 1200 stones in approximately 68 seconds. The once-off per silhouette set preprocessing takes approximately 50 seconds. Computing dissimilarity between shape distributions takes approximately 0.7 seconds and pose optimisation takes 17 seconds (of which 15 seconds is spent considering matches and 2 seconds on mismatches).

Varying Moments and Shape Approximation Methods Used

A set of experiments was carried out to determine the effects of the number of moments used to form initial estimates and the shape approximation method used.

Using only first order moments (moments up to order 1) means using only the centroids of the shape approximation (VEMH, visual hull, or CDRH) to form the positional component of initial pose; the rotational component is random.

Using first and second order moments (moments up to order 2) makes use of the principal axes of the shape approximation for the first four initial pose estimates. The four possible alignments of the principal axes are considered in random order.

Using first, second, and third order moments (moments up to order 3) uses the third order moments to order the four possible alignments of the principal axes as described in Chapter 6.

Table 9.2 shows the mean time over 30 trials spent on pose optimisation. The same starting point selection and shape approximation methods used for testing were also used for training in each case. The results indicate that the VEMH produces shorter running times than the visual hull and the CDRH. Using more moments for initial pose estimates reduces running times.

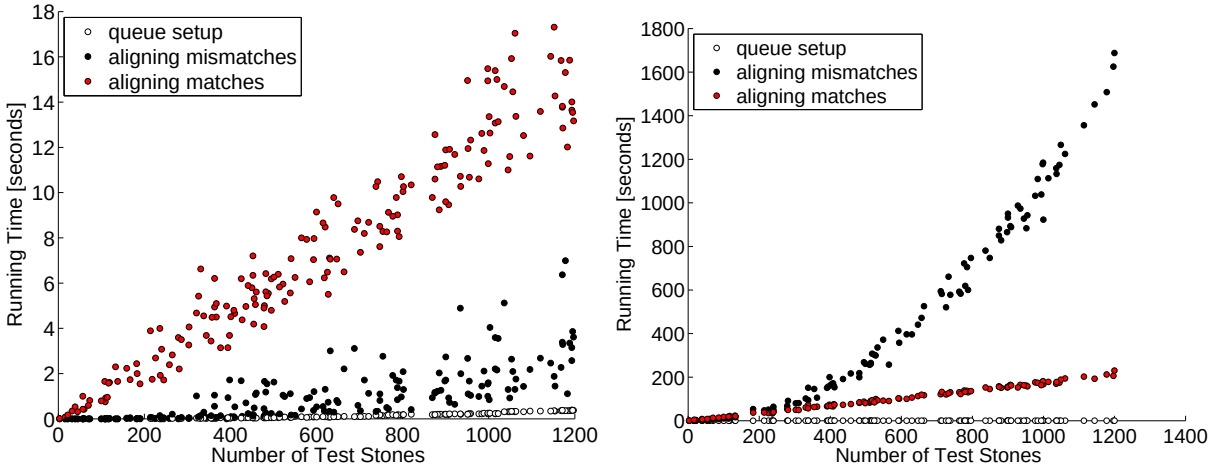
	VEMH	VH	CDRH
moments up to order 1	273.5 (208.7, 64.8)	515.6 (207.4, 308.3)	1796.6 (210.4, 1586.2)
moments up to order 2	30.4 (22.0, 8.4)	65.3 (27.4, 37.8)	297.1 (39.9, 257.2)
moments up to order 3	17.4 (15.0, 2.4)	38.7 (21.4, 17.2)	198.0 (35.6, 162.5)

Table 9.2: Mean running time (in seconds) spent on applying pose optimisation for batch matching of 1200 silhouette sets of uncut gemstones across two runs. Times spent on matches and on mismatches are shown in brackets.

Running Time as a Function of Number of Stones

The next set of experiments investigates how running time is affected by varying the number of stones. Random subsets of up to 1200 stones were selected as test sets.

Figure 9.2a shows a plot of number of stones versus running time using the VEMH and moments up to order 3 for determining starting points.



(a) (b)

Figure 9.2: Running time for batch matching different numbers of stones: (a) using the VEMH and moments up to order 3, (b) using the CDRH and moments up to order 1. The running time consists of time spent setting up the priority queue and applying pose optimisation to matches and to mismatches.

Setting up the priority queue takes only a small amount of time, yet populating the priority queue is of $O(m \log m)$ complexity for m elements. Since there are $m = n^2$ elements for n stones, the time complexity is $O(n^2 \log n^2)$; setting up the priority queue will become the most time consuming component for sufficiently large n. Applying pose optimisation to match pairs takes time proportional to the number of stones. However, for the values tested here it forms the largest component of the running time. Although the time spent on

mismatches is $O(n^2)$, for values of n tested, the running time is small. For values of n up to 1200, the running time is therefore approximately proportional to the number of stones.

The experiment was repeated using the CDRH and moments up to order 1 (only random rotations were used). This was done to observe the quadratic dependence of the time spent optimising mismatch pairs on the number of stones. The results are shown in Figure 9.2b. The quadratic dependence is more apparent than in Figure 9.2a. A larger proportion of running time was spent considering mismatch pairs than match pairs for larger numbers of stones.

Using Downsampled Input Images

Image downsampling was used to investigate the behaviour of batch matching in cases in which image noise is high enough that some ET errors across a silhouette set fall on the wrong side of the threshold. Since the proposed batch matching algorithm applies pose optimisation to pairs until all matches are found, the algorithm will fail to terminate if there are insufficient pairs with ET errors below the threshold. A limit on the time spent on pose optimisation must therefore be imposed to force termination.

Table 9.3 shows the error rates achieved for different degrees of downsampling and for different time limits. (An error is incurred if a silhouette set is matched to the wrong silhouette set or is not matched at all; the

	Time Limit [seconds]					
	5	10	20	40	80	160
original resolution	54.4%	22.0%	0.5%	0.0%	0.0%	0.0%
2×2 binning	57.1%	26.3%	0.5%	0.0%	0.0%	0.0%
4×4 binning	53.2%	14.8%	0.5%	0.5%	0.4%	0.4%
8×8 binning	41.8%	22.9%	9.6%	8.7%	8.7%	8.8%
16×16 binning	61.6%	39.2%	26.8%	14.4%	12.6%	12.7%
32×32 binning	76.8%	65.6%	60.8%	60.1%	60.1%	60.1%
64×64 binning	96.0%	95.5%	95.5%	95.5%	95.5%	95.5%

Table 9.3: Mean error rates over 30 trials for batch matching two runs of 1200 stones with a time limit imposed on the running time spent on pose optimisation. Results are shown for different levels of downsampling (pixel binning). Each error rate corresponds to batch matching of two runs of 1200 silhouette sets of uncut gemstones. Images were segmented using the subpixel resolution method described in Appendix A.

error rate is the number of errors divided by the number of stones.) There is little reduction in the error rate between 40 and 80 seconds, indicating that further matches are unlikely to be found. At levels of downsampling greater than 2×2 binning, the silhouette sets are not all correctly matched up for even the largest time limit. The error rate increases as the degree of image downsampling is increased.

The approach of imposing a time limit may be useful for cases where image resolution is poor. Image resolution may be insufficient for all match and mismatch errors to be on opposite sides of the ET error threshold; however, a 100% correct matching may not be a necessity. This situation can occur in cases where

one is interested in estimating statistical shape properties of a batch of stones using merged silhouette sets. If a small number of silhouettes is not matched, or is incorrectly matched, this may have negligible effect on the shape property estimates, especially since mismatched pairs will tend to be of similar shape.

9.3.3 Batch Matching with Caliper Diameter Distributions

An experiment was carried out to investigate the error rates and running times achievable using only the likelihood ratios derived from EMDs between caliper diameter distributions (i.e., not using ET-based pose optimisation.) The same likelihood ratio values that were computed for the experiments described in Section 9.3.2 were used as input.

The maximum likelihood permutation is the permutation that results in the highest product of likelihood ratios. To compute the permutation, the logarithm of likelihood ratios is used, so that the sum can be maximised, rather than the product. Finding the permutation of a square matrix that minimises summed cost is a well-known combinatorial optimisation problem that can be solved using the Hungarian Method [23]. A Matlab implementation of the Hungarian Method (provided by Niclas Borlin of Umeå University, Sweden) was used to determine the permutation that maximises the sum of the logarithm of likelihood ratios.

Results are shown in Figure 9.3. Each data point corresponds to an experiment in which 10 runs of 223

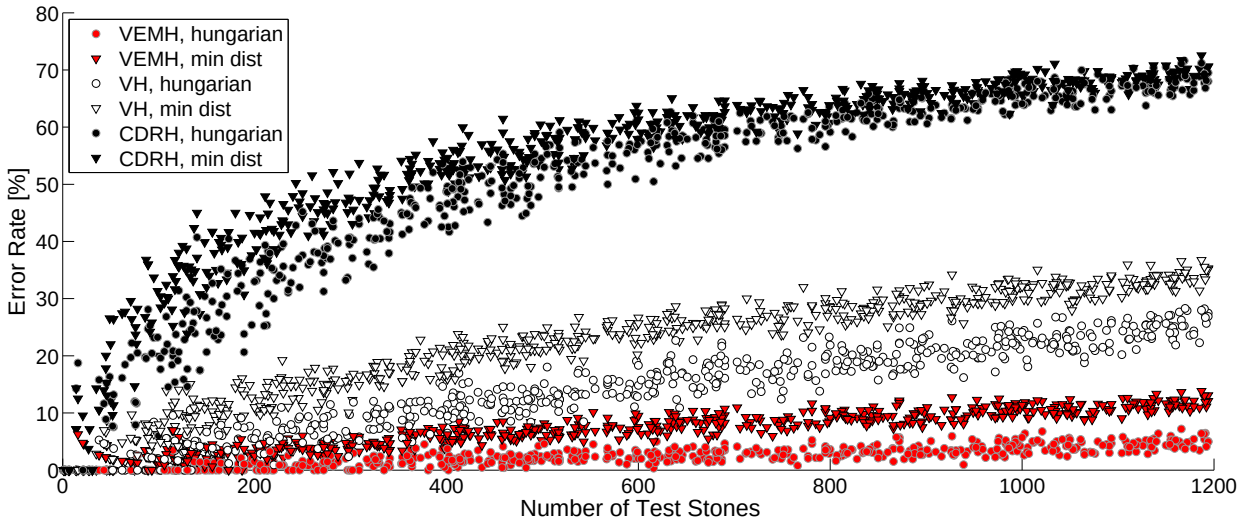


Figure 9.3: Plot of number of test stones versus error rate for batch matching based on caliper distribution using different methods for approximating stone shape. Two methods are used for determining the permutation matrix: the Hungarian Method, and selecting the row with the minimum EMD for each column of the permutation matrix.

randomly selected stones are used as a training set and the test set of stones is randomly selected from the remaining 1200 stones. The two runs used as the test set for each data set were randomly selected from the 10 available runs. For each training and test set, separate results were computed using the VEMH, visual

hull and CDRH for shape approximation. The match permutation was computed using both the Hungarian Method, and a simpler minimum distance method that selects the row with the minimum EMD for each column of the permutation matrix.

The results indicate that the batch matching using only EMDs cannot be carried out without error for the data sets considered. The VEMH outperforms the visual hull, which in turn outperforms the CDRH for shape approximation. The Hungarian Method outperforms the minimum distance method for forming the permutation matrix from a square matrix of EMD values. For batch matching of 1200 stones, the most accurate approach (VEMH for shape approximation and the Hungarian Method for computing the permutation) achieves an error rate of approximately 5%.

These experiments demonstrate that the information contained in the EMD values is an important aid to the batch matching process, but is alone insufficient. ET-based pose optimisation must also be used for matching that is both efficient and correct.

9.4 Summary

An algorithm has been designed and implemented for efficiently matching two runs of silhouette sets of the same batch of stones. Various approaches were combined to ensure efficiency:

1. Likelihood ratios based on rapidly computed EMD values between estimated caliper distributions are used to identify the pair (of those still under consideration) that is most likely a match.
2. ET-based optimisation is applied to the most likely match pair from a single starting point before updating the likelihood ratio for the pair if the pose optimisation fails.
3. If a match is found (alignment with sufficiently low ET error across the two silhouette sets), then pairs that are implied to be mismatches are removed from consideration.
4. Moments of 3D approximations to the stone computed from pairs are used to select initial pose estimates most likely to lead to correct alignment of the two silhouette sets.

On a test set of 1200 uncut gemstones, pairs of runs of six-view silhouette sets are matched in approximately 18 seconds on a 3.2 GHz Pentium 4 machine. The once-off per silhouette set processing takes approximately 50 seconds; the computations are sufficiently fast to be computed online as the stones are passed through the six-camera setup. This represents a substantial improvement on a naive approach where alignment-based matching is applied to all pairs (such an approach would take weeks to complete). The proposed approach is also superior to the naive approach in that the number of starting points to consider for aligning each pair need not be decided in advance.

For correct matching, the proposed method requires the minimum ET error for all match pairs to be below a pre-specified threshold (determined with a training set), and the error for all mismatch pairs must be above this threshold. Although this is the case for the data sets and camera configuration considered, this is not guaranteed to hold. If the criterion fails to hold, then the algorithm may fail to terminate. To force termination, a limit can be imposed on the time spent on pose optimisation. The effects of applying the algorithm to cases in which silhouette set quality is insufficient to meet the criterion has been demonstrated by using downsampled versions of the original images. As the time limit is increased, improvements in the error rate become negligible. As the degree of downsampling is increased, the error rate increases.

Experiments that use the Hungarian Method to estimate the match permutations using likelihood ratios based on EMDs between signatures, and not using ET-based alignment, produced errors (a correct classification rate of approximately 95% is achieved for matching 1200 silhouette sets across two runs). This justifies combining the signature-based matching with alignment-based matching to achieve results that are both correct and efficiently computable.

Chapter 10

Comparing Silhouette-Based Sizing with Sieving

10.1 Introduction

Particle shape analysts are interested in (1) emulating sieving with silhouette-based methods, (2) quantifying the repeatability of silhouette-based sieve emulation, and (3) investigating the effect of individual particle shape on the sieve aperture through which the particles pass [42, 109]. This chapter describes an experiment which uses the methods developed in this thesis to address all three of these issues.

The repeatability of sieve sizing cannot be evaluated by sieving particles individually. The sieve bin that each particle ultimately lands in is a function not only of particle shape, but also the length of time over which the sieves are shaken, and the presence of other particles in the sieves.

Knowing which sieve bins each particle lands in over multiple runs of batch sieving provides (1) an understanding of the shape characteristics that determine bin classification and bin classification variability, and (2) a more accurate quantification of repeatability than if histograms alone were considered. By appropriately quantifying the repeatability of sieving, it can be directly compared with silhouette-based sizing methods. Demonstrating that silhouette-based methods are at least as repeatable as sieving is an important step in having such methods accepted by particle shape analysts as an alternative to sieve sizing.

The minimum enclosing cylinder of a silhouette-based 3D approximation to the stone shape (both the VEMH and the visual hull are tested) is used for sieve emulation. This approach is based on the assumption that the minimum enclosing cylinder of a stone provides a good approximation of the smallest circular sieve aperture through which the stone may pass. By comparing the silhouette-based estimate of minimum cylinder diameter with the sieve bins in which each stone is actually found to land up in, the accuracy of the silhouette-based sieve emulation can be quantified.

The experiment described in this chapter was carried out as follows. A data set of 494 garnets was sieved fifteen times using a stack of sieves with five bins. (The 494 stones are illustrated using refined visual hull models in Appendix C on page 227.) The five sieve bins are separated by circular sieve apertures with diameters of 4.521 mm, 5.410 mm, 5.740 mm, and 6.350 mm. For each of the fifteen sieving runs, the garnets were manually sieved for 30 seconds. After each run of sieving, the garnets were passed through the six-camera setup in five sub-batches according to the sieve bin in which they landed. This means that the sieve bin corresponding to each six-view silhouette set is known. By matching the silhouette sets that correspond to the same stones across the fifteen runs, the bin in which each stone landed for each run is determined.

The experiment allows the performance of the matching procedure to be evaluated too. Although the stone identity associated with each silhouette set is not known in advance, consistency constraints across multiple runs can be used to evaluate matching performance. For instance, if A matches B, and B matches C, then A must match C.

10.2 Batch Matching

For the purposes of matching the stones across runs, batch matching was applied to batches of silhouette sets from each run and the run's immediate successor. Preprocessing (computing signatures and moments from the raw image sets) requires approximately 20 ms per silhouette set. Although for this experiment preprocessing was carried out offline, it is sufficiently fast to be carried out online as the stones pass through the feeder at a rate of approximately 10 stones per second. After preprocessing, batch matching across two runs of the 494 stones requires approximately three seconds of processing time. (This is faster than matching the same number of gemstones because the garnets are less compact and therefore alignment tends to require fewer optimisations.) The batch matching is therefore substantially faster than sieving the stones and feeding them through the six-camera setup, and can clearly be considered to be sufficiently fast for practical experimental purposes.

Batch matching was applied to all run pairings to check consistency. There are $\binom{15}{2} = 105$ pairs of runs that are formed from the 15 runs. The data set of 246 garnets was used as training data to determine parameters for the experiments described in this chapter.

A necessary condition for correct matching is that the matching results are consistent across runs. If a silhouette set from Run A matches one from Run B and one from Run C, then the silhouette sets from Run B and Run C must match each other. By adding up the number of cases in which triplets of pairwise matches are consistent, a measure of consistency can be made. There are $\binom{15}{3} = 455$ triplets of runs and 494 cases in each run, yielding a total of $455 \times 494 = 224\,770$ triplets.

All 224 770 triplets were found to be consistent for the batch matching. This result was compared with other simpler approaches to matching. Table 10.1 shows the results for different matching methods that were tested. The first column of the table (Minimum Dissimilarity) shows results for matching using the minimum

Matching Method	Minimum Dissimilarity	Square Assignment
caliper signatures	94.0%	98.8%
caliper, three principal	76.0%	84.0%
VH min cylinder diameter	62.9%	19.8%
VEMH min cylinder diameter	63.0%	24.5%
VH nonconvex volume	63.1%	26.7%
VH convex volume	63.5%	25.5%
VEMH convex volume	63.4%	27.7%

Table 10.1: Percentage of consistent triplets of pairwise matches computed using different matching methods.

dissimilarity case in the other run as the match. The second column (Square Assignment) shows results for matching using square assignment, which selects the permutation that results in the maximum sum of log likelihood ratios (described in Section 9.3.3).

The first row of the table shows the results of matching using dissimilarity based on the caliper signatures described in Chapter 8. The second row of the table uses results based on dissimilarity defined as the Euclidean distance between three caliper diameters of the VEMH. The remaining rows of the table define dissimilarity using the differences between minimum enclosing cylinder diameters and difference between volumes for visual hulls and VEMHs.

The table indicates that, unlike the proposed batch matching method, none of the other methods is perfectly consistent. Consistency is a necessary but not sufficient condition for correct matching. If A is similar to B, and A is similar to C, then B is likely to be similar to A even if the measure of similarity is inaccurate. For this reason, the first column shows consistency values of over 60% for relatively poor approaches to matching, such as choosing the case in which the difference between minimum cylinder diameters of the visual hull is a minimum. (A computer simulation in which 15 runs of 494 random measurement values were drawn from a uniform distribution, and minimum dissimilarity was used for matching, resulted in 62.5% of triplets being consistent.)

Table 10.2 shows the percentage of cases that are correctly matched for the various methods. These are computed on the assumption that the proposed batch matching method produces correct results. The table indicates that matching based on a single shape property performs poorly, with VEMH-based estimates outperforming visual hull-based estimates. Note that it is possible to have a greater percentage of correct matches than consistent triplets: the $105 \times 494 = 51\,870$ matched silhouette set pairs do not correspond directly to the $455 \times 494 = 224\,770$ triplets.

Matching Method	Minimum Dissimilarity	Square Assignment
caliper signature	96.6%	99.4%
caliper, three principal	79.1%	87.7%
VH min cylinder diameter	3.6%	3.2%
VEMH min cylinder diameter	5.7%	5.3%
VH nonconvex volume	6.5%	6.0%
VH convex volume	6.8%	6.4%
VEMH convex volume	7.8%	7.3%

Table 10.2: Percentage correct matches as classified by the ET-based batch matching.

10.3 Silhouette-Based Sieve Emulation

This section investigates sieve emulation using the minimum enclosing cylinder of a 3D approximation of the stone. The same silhouette sets that are used for batch matching are used to approximate 3D shape.

A analogous approach is used by Fernlund et al. [43] who provide a method for emulating square-aperture sieves. They compute the minimum enclosing square from all available silhouettes of each stone. This is used to approximate the smallest infinite-length prism with a square cross section that encloses the stone.

10.3.1 Computing the Minimum Enclosing Cylinder

The minimum enclosing cylinder of a 3D point set is the smallest diameter cylinder of infinite length that completely encloses the points. Various methods for estimating the minimum cylinder have been developed in the field of computational geometry, but implementing these methods is non-trivial [24, 116]. The method proposed here instead uses a conjugate-gradient minimisation algorithm to minimise the cylinder radius from many starting points corresponding to different directions. An efficient minimum enclosing circle algorithm [135] is used to compute the minimum cylinder radius for each direction, by projecting all points onto a plane that is perpendicular to the direction of the cylinder axis. The derivative of the cost function is required by the conjugate-gradient minimisation algorithm. This was calculated using an azimuth-elevation representation for directions. To computed the partial derivatives, only the support points of the circle on the plane need be considered. Since a circle is supported by either two or three points (barring cases in which an infinitesimal perturbation of the points changes the number of support points), both cases need to be formulated. The Matlab Symbolic Toolbox was used to compute a solution, which was verified using a forward difference approximation. (The resulting C code for the derivative computation is tens of thousands of lines long.) One hemisphere of a subdivided icosahedron is used to create direction samples. Some computer simulations were carried out to determine a set of parameters (number of direction starting points, number of descent iterations, number of optimisations from n best starting points) with desirable speed-accuracy tradeoff characteristics.

10.3.2 Experimental Results

Minimum enclosing cylinders were computed from the 15 runs of 494 silhouette sets using the visual hull and the VEMH to approximate the stone shape. The minimum cylinder was expected to provide a reasonable estimate of the minimum sieve aperture that the corresponding stone can pass through. This provides a means for predicting the sieve bin that the stone would land in from its silhouette set. The minimum cylinder is used as an approximate means for predicting sieve bins; certain stones (e.g. banana-shaped stones) may pass through sieve diameters that are smaller than their minimum cylinder diameter. A similar observation is made by Rao [109] also in the context of silhouette-based particle sizing: "[T]here is a chance that the particle can weave, wiggle and make its way through... under the vigorous sieve shaking process..." Although Rao makes use of square-aperture sieves, the observation is equally valid for the circular-aperture sieves used here.

The minimum enclosing cylinder diameters were used to classify each silhouette set into one of five bins using the sieve aperture diameters as bin boundaries. The proportions of cases in each bin over all 15 runs is given in Table 10.3. These values give an indication of the extent to which the silhouette-based methods can be used to emulate sieving.

Bin No.	1	2	3	4	5
Sieving	14.1%	32.6%	19.1%	16.9%	17.2%
Min Cylinder VH	7.5%	31.8%	13.3%	25.4%	21.9%
Min Cylinder VEMH	10.6%	32.5%	15.1%	22.7%	19.2%

Table 10.3: Mean proportion of cases in each of the five bins

Histograms for the sieving and silhouette-based emulations are shown in Figure 10.1. The histograms give an indication of the extent to which the silhouette-based methods can emulate the sieving process, as well as an indication of the repeatability of the different sizing methods from run to run.

To investigate the extent to which the minimum cylinder diameter of a stone is a good estimate of the smallest sieve aperture that the stone can pass through, 90-view visual hulls were formed for each stone by merging silhouette sets (as described in Chapter 6). The minimum cylinders of the 90-view visual hulls are assumed to be good estimates of the minimum cylinders of the corresponding stones. Each stone's minimum sieve bin over the 15 runs was used as an estimate of the smallest of the five bins that the stone could land in, i.e., it was assumed that if the stone *could* pass through a sieve aperture, then it *did* pass through on at least one of the 15 runs. Figure 10.2 shows histograms of the minimum cylinder diameters for stones corresponding to each of the five bins. Vertical lines indicate the locations of the four bin boundary aperture diameters. Since the minimum cylinder diameter cannot be smaller than the smallest circular aperture through which a stone can pass, the histograms are all expected to lie to the right of the lower bin boundary. Figure 10.2 shows that this is indeed the case: the histograms lie to the right of the vertical lines that represent the sieve aperture diameters. This means that all stones passed through all apertures wider than their minimum cylinders on at

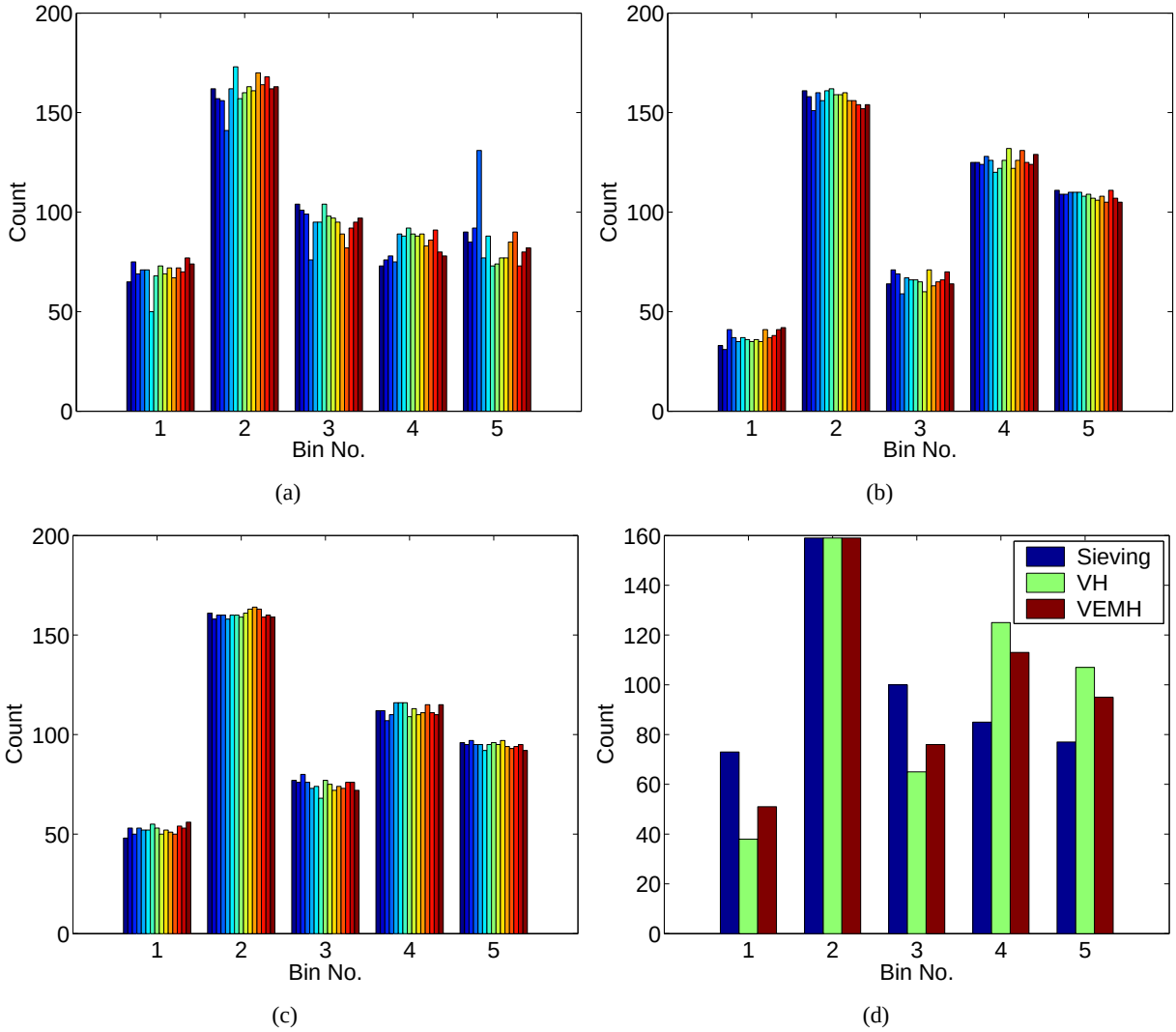
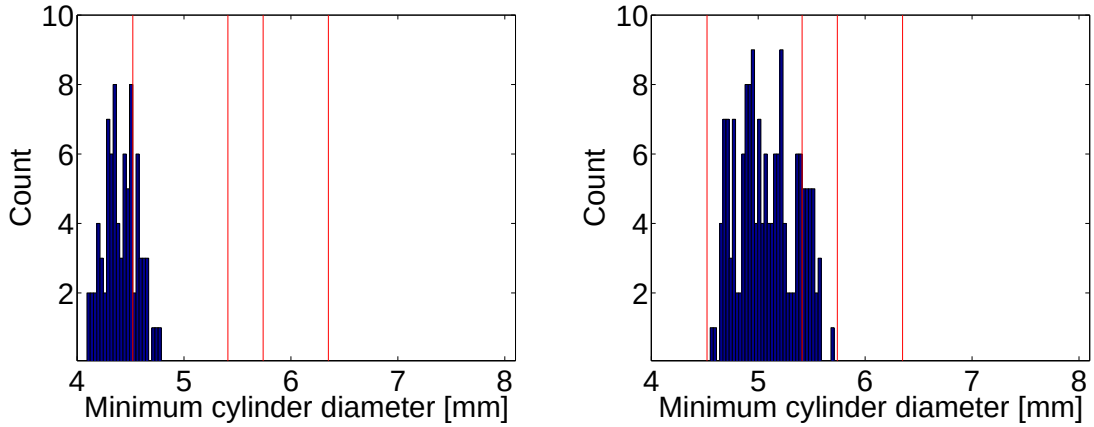
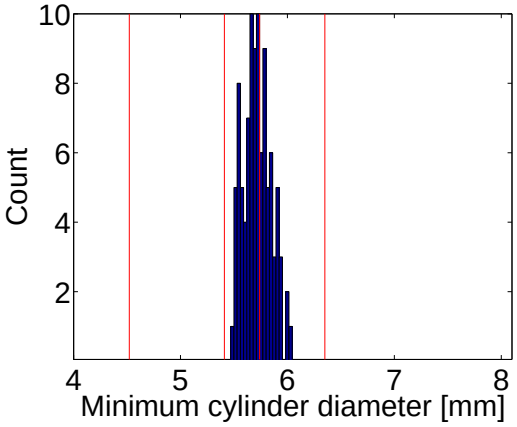


Figure 10.1: Histograms for (a) sieving, (b) minimum cylinder of the visual hull, (c) minimum cylinder of the VEMH, and (d) mean bin counts for 15 runs of the three sizing methods. Each run is represented by a different colour bar for histograms (a)–(c).

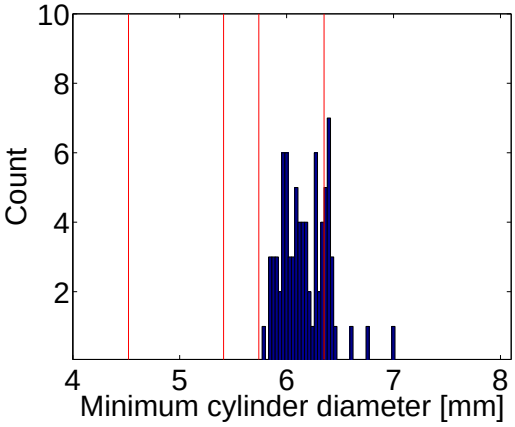


(a) Bin 1

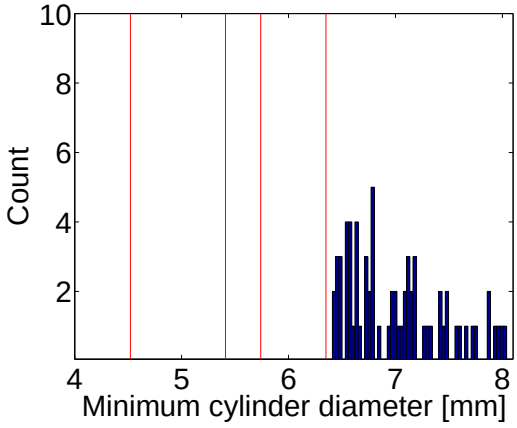
(b) Bin 2



(c) Bin 3



(d) Bin 4



(e) Bin 5

Figure 10.2: Histograms of minimum enclosing cylinder diameters for 90-view visual hulls of garnets whose minimum sieve bin is (a) Bin 1, (b) Bin 2, (c) Bin 3, (d) Bin 4, (e) Bin 5. Vertical lines indicate the sieve aperture diameters.

least one of the 15 runs. Parts of the histograms that overlap the upper bin boundaries correspond to cases in which the stones have passed through an aperture smaller than their minimum cylinder. The figure indicates that there are cases of stones passing through apertures up to 10% smaller than the minimum enclosing cylinder. Note that a stone that passes through an aperture 10% smaller than its minimum enclosing cylinder may pass through an aperture even smaller than this. This means that the ratio of minimum cylinder diameter to smallest possible sieve aperture diameter (i.e., considering sieves of any diameter rather than the four used in this experiment) is likely to be larger than 1/(1-10%).

Figure 10.3 illustrates the fives stones whose minimum sieve bin is smaller than the minimum cylinder diameter by the largest amount. These stones must exhibit some degree of concavity, since convex stones

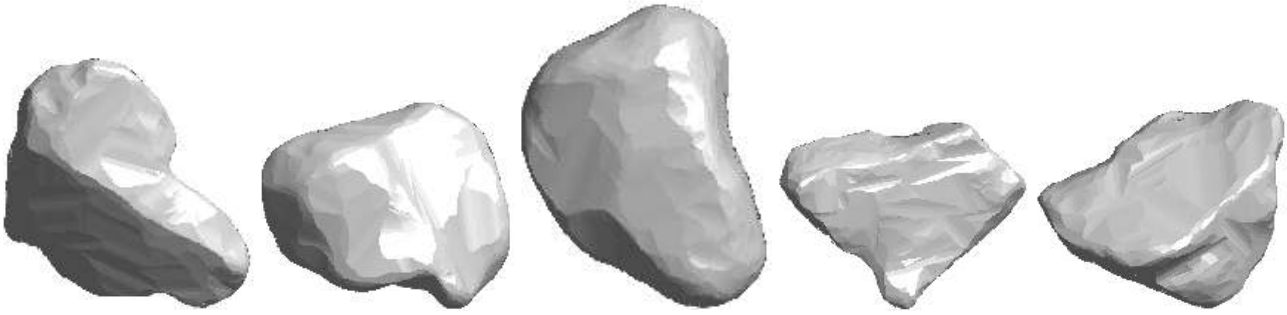


Figure 10.3: Refined visual hull models of stones whose minimum sieve bin is smaller than the minimum cylinder diameter by the largest amount.

cannot pass through a sieve aperture of smaller diameter than the minimum enclosing cylinder. It is visually apparent that the concavities on these stones allow the stone to pass through a sieve aperture smaller than its minimum enclosing cylinder.

No attempt was made to improve the estimate of the smallest sieve aperture through which a stone can pass by accounting for possible changes in the direction of motion as the stone passes through a sieve aperture. However, it is interesting to note that for a convex-shaped aperture (such as the circular or square apertures used in practice), the line hull of any shape that can pass through the aperture (with possible changing direction of motion) can pass through the aperture too. This is because synclastic concavities (such as a dimple in a golf ball) do not affect whether a shape can pass though a convex-shaped aperture. It is therefore possible, in principle, to determine whether or not a 3D solid can pass through a convex-shaped aperture, by considering only its silhouettes from all viewpoints.

10.4 Comparing Histogram Repeatability

Particle shape analysts have historically made use of histograms, which are the natural output of sieving to quantify the size characteristics of a batch of particles. It has been argued that particle volume measurements are often preferable to sieve size measurements for the purpose of characterising particle size [133]. To

switch from sieving to silhouette-based volume estimates for characterizing size, particle shape analysts require that histograms of volume estimates are at least as repeatable from run to run as histograms derived from sieving.

This section considers how the problem of comparing histogram repeatability can be meaningfully framed, and then provides the results of an experiment that indicates that both silhouette-based sieve emulations and volume estimates provide more repeatable histograms than the sieving runs carried out for the data set of 494 garnets.

10.4.1 Method for Comparing Histogram Repeatability

Summing Bin Count Variances: ΣVAR

The variation of histogram bin counts from run to run provides a means for computing repeatability. For a perfectly repeatable system, the bin counts will not vary from run to run. The repeatability of two measurement systems can be compared using the sum of bin count variances over multiple runs of histograms produced by the two systems using the same sample of stones. The sum of bin count variances, ΣVAR, for r histograms produced by a measurement system with n bins is given by

$$\Sigma\text{VAR} = \sum_{j=1}^{n} \text{VAR}(\mathbf{b}_j), \tag{10.1}$$

where $\mathbf{b}_j$ is a vector of length r containing the counts of the jth bin, and VAR gives sample variance. Lower ΣVAR indicates greater repeatability.

Individual Stone Contributions

If ΣVAR is to be determined by binning individual measurements, then different combinations of histograms are possible. For instance, the first histogram may use the first measurement of the first stone and the first measurement of the second stone, or it may use the second measurement of the first stone and the first measurement of the second stone. All possible combinations are equally valid, since each stone measurement is made independently of all others. The variation on the ΣVAR statistic due to the specific combination of measurements used to form the histograms can be reduced without introducing bias by summing the bin variances for the measurements corresponding to each stone individually.

For n measurements of a stone, the contribution of a bin to the ΣVAR statistic is

$$\text{bin variance contribution} = \frac{kn - k^2}{n(n-1)}, \tag{10.2}$$

where k is the number of times the measurement falls into the bin. The bin variances for all bins and all stones must be added to form the ΣVAR statistic.

Naive Bin Boundary Specification

To compare the repeatability of histograms formed from silhouette-based estimates of properties such as volume with sieve histograms, the numbers of bins must be equal, and the probability of assigning a case to corresponding bins must be the same. (If this were not the case then bin boundaries could be chosen to create arbitrarily low ΣVAR values.)

In order to determine the bin boundaries for the silhouette-based estimates (for which the individual measurements are available), the total proportion of measurements in each of the sieve classes must be calculated. The bin boundaries must be positioned so that the same proportion of the total measurements are classified into the corresponding classes. It is not, however, a straightforward matter of ensuring that the same proportion of measurements fall into each class.

Consider Figure 10.4. Each of four stones, represented by the +, ×, ⋆ and ○ symbols, has been measured

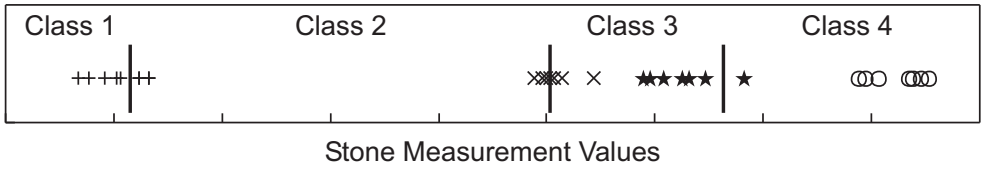


Figure 10.4: Class boundaries cannot be determined using the stone whose measurements are to be classified with the boundaries.

seven times. If the histograms with which the measurement system is to be compared have class probabilities of 5/28, 6/28, 9/28 and 8/28, then the class boundaries shown as vertical lines in the figure would divide the measurements so that the proportions correspond to the other measurement system. This method introduces a bias, since no matter how tightly the measurements from each stone cluster, measurements will be assigned to different bins resulting in an apparently poor repeatability statistic for a repeatable system.

Leave-One-Out Bin Boundary Specification

To classify measurements without introducing this bias, a leave-one-out approach is used. The measurements corresponding to each stone are classified individually, using the measurements of the remaining stones. The procedure is illustrated with an example. Consider an experiment in which five stones are each measured eight times using a particular device. If the device is to be compared with a system that outputs histograms with bin probabilities of 25%, 45% and 30% for bins #1, #2 and #3 respectively, then measurements can be

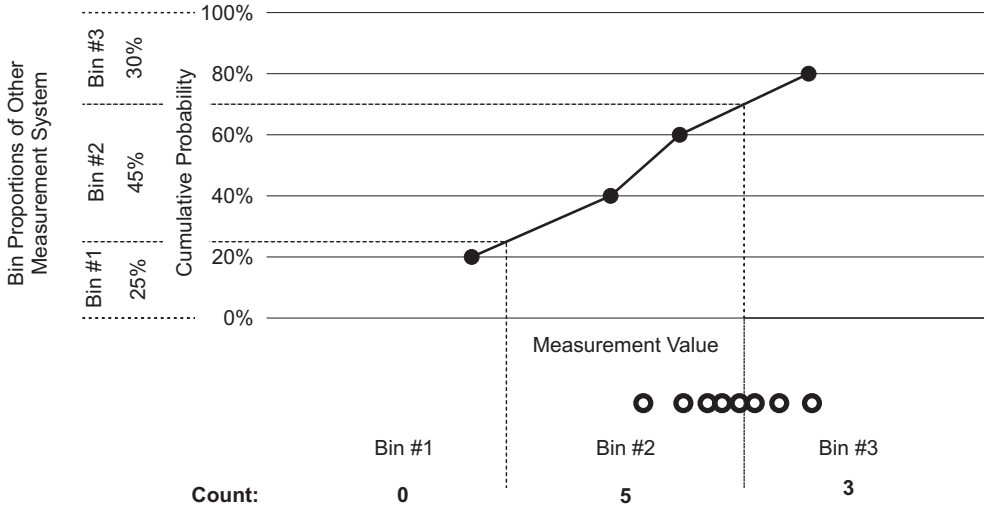


Figure 10.5: An example of classifying measurements into bins based on the measurements of the remaining stones and the mean histogram of the of the other measurement system.

classified according to the bin boundaries as determined in Figure 10.5. In this example, the eight measurements for a particular stone are represented by the open circles shown in Figure 10.5. The bin boundaries for determining which bin each of these eight measurements falls into are determined from the sets of repeated measurements for the four remaining stones. The mean of the eight measurements for each of the four stones is computed (shown as black dots) and define the cumulative probability distribution. The bin boundaries are then determined from this cumulative distribution and the eight measurements for the single stone being binned are classified into the corresponding bins. Five of the measurements are binned into Bin #2 and three of the measurements are binned into Bin #3. The histogram bin counts for each run (or measurement) are shown in Table 10.4. The variance for each bin can be calculated using Equation 10.2. This particular stone contributes a total of $30/56$ to the ΣVAR statistic.

	Bin #1	Bin #2	Bin #3
Run #1	0	1	0
Run #2	0	0	1
Run #3	0	1	0
Run #4	0	0	1
Run #5	0	1	0
Run #6	0	1	0
Run #7	0	1	0
Run #8	0	0	1
VAR	0	$15/56$	$15/56$

Table 10.4: Histogram bin counts for single stone and corresponding contributions to ΣVAR statistic

Note that for n stones, extrapolation is required to define bin boundaries for cumulative bin probabilities below $1/n$ and above $(n-1)/n$. For a large number of stones, this situation is unlikely to occur. For cases

in which the situation does occur, the contribution of the small number of measurements that do fall outside the range can be discounted and the ΣVAR weighted accordingly.

10.4.2 Experimental Results

The ΣVAR statistic can be estimated more accurately from individual measurements than directly from histograms. However, this is under the assumption that the individual measurements are independent of the run. The assumption may not be valid for sieving, with individual measurements being affected by the sieving vigour for each run. Nevertheless, the ΣVAR statistic provides a useful means for comparing silhouette-based measurements with the inherently quantised sieving measurements in the sense of individual measurement repeatability. In addition, if a silhouette-based method can be shown to be more repeatable in the sense of individual measurement variability, then it must also be more repeatable in the sense of direct histogram variability, since the run-dependent influence cannot decrease variability.

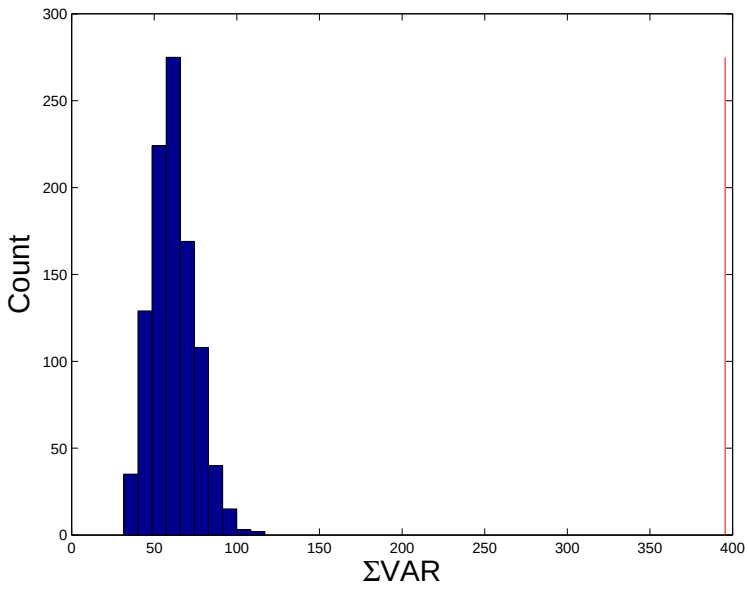


Figure 10.6: Histograms of ΣVAR values computed directly from sieve histograms with 1000 trials of random permutations of the 15 bin values for each stone. The ΣVAR value of 393.0 computed from the original data is indicated with a vertical line.

Figure 10.6 illustrates the effect of randomly permuting the 15 bin values for each run on the computed ΣVAR statistic. This removes the run-dependent variability component. Each of the 1000 trials produced a ΣVAR value much lower than that computed from the original permutation, providing strong evidence that there is a large degree of run-dependence on the bin values.

To test the validity of the implementation of the method for estimating ΣVAR from individual measurements, a computer simulation using synthetic data was set up. Sieve bins were computed for 15 runs of 494 stones, using proportions of bin occurrences for the real data to derive the distributions from which random values were drawn. The experiment was repeated 250 000 times, with the ΣVAR statistic being computed directly

from histograms and also from individual measurements for each trial. The mean and standard deviations of these values are given in the first and second rows of Table 10.5. Note that the mean values are similar for the

Method	Runs	Trials	Mean	STD
ΣVAR from histograms	15	250 000	57.426	12.435
ΣVAR individual measurements	15	250 000	57.385	1.639
ΣVAR from histograms	100 000	10	57.472	0.140
ΣVAR individual measurements	100 000	10	57.389	0.017

Table 10.5: Results of a computer simulation in which the ΣVAR statistic was computed directly from histograms and from individual measurements for 494 cases. Values were generated from probabilities determined by the proportions of bin occurrences for each of the 494 garnets over the 15 runs of sieving.

two methods, but the values computed from the histogram show a much larger spread than those computed from the individual measurements. This indicates that a better estimate of ΣVAR is obtained using individual measurements. To ensure that the correct quantity is being measured, the ΣVAR statistic was computed for a large number of runs. The results of 10 trials of 100 000 runs are given in the third and fourth rows of the tables The results indicate that estimating ΣVAR from a small number of runs (15 runs) does not introduce substantial bias with either the histogram or individual measurement methods.

Comparing the silhouette-based measurements with sieving measurements requires the sieving measurements to be binned. Bin probabilities must be the same as the sieve bin probabilities for a meaningful comparison.

A computer simulation was carried out to test the validity of the implementation. A normal distribution was created from which 494 sample values were drawn. Normally distributed noise of fixed standard deviation was then added to create 15 noisy measurements for each sample value. The samples were then binned into five bins using four bin boundaries. These data represent the sieve measurements. The ΣVAR statistic was computed from these measurements using the individual measurement method. A new set of measurements was then drawn from the same distributions. These represent silhouette-based measurements with the same inherent repeatability as the simulated sieve measurements. The ΣVAR statistic was computed for these measurements using the simulated sieve data to assign the data to bins using the leave-one-out approach. Note that the bin boundaries are not used, and in general will be meaningless when comparing the repeatability of two measurement systems that may be measuring different stone properties (e.g., mass, volume, hardness, electrical conductivity). In addition to the data generated from the same distributions, data were also generated using measurement noise distributions with 1.1 and 0.9 times the original standard deviations. The ΣVAR statistic should indicate that these measurements are less repeatable and more repeatable respectively. Mean and standard deviations of ΣVAR values for 500 trials are shown in Table 10.6. The results indicate that the leave-one-out method produces a ΣVAR statistic that is in close agreement with the directly computed ΣVAR statistic from an equally repeatable measurement system. The ΣVAR values also correctly reflect the lower and higher repeatability of the simulations of the two other measurement systems with different noise characteristics.

Data	Mean	STD
Original Binned Data	59.92	3.88
Data with same measurement STD	59.59	4.34
Data with 1.1× measurement STD	65.44	4.62
Data with 0.9× measurement STD	53.93	4.02

Table 10.6: Results of a computer simulation in which the ΣVAR statistic was computed by binning individual measurements using the leave-one-out approach. The mean and standard deviations for 500 trials are shown. Size and noise variation values were based on minimum cylinder diameter values computed from the real data set. The four sieve aperture diameters were used as bin boundaries.

ΣVAR statistics were computed for the sieving data and various silhouette-based estimates of shape properties. Results are presented in Table 10.7. The 95% confidence intervals were computed using the bootstrap

Measurement	ΣVAR	95% Conf. Int.	
		lower bound	upper bound
sieving	61.5	53.9	69.6
VH min cylinder diameter	64.0	53.7	77.5
VEMH min cylinder diameter	35.3	27.1	46.4
VH nonconvex volume	30.4	25.6	43.6
VH convex volume	29.8	24.8	41.5
VEMH convex volume	24.1	20.0	36.5

Table 10.7: ΣVAR statistics for sieving and various shape features measured from silhouette sets.

percentile method [36] with 2000 bootstrap samples per case. Note that the sieving ΣVAR values that are computed from individual measurements are substantially lower than the values computed directly from histograms (see Figure 10.6 in which the mean value of the histograms will tend towards the values in Table 10.7 as the number of trials is increased).

Although the sieving ΣVAR values which are computed from individual measurements would provide an underestimate of the actual sum of bin variances one would obtain over repeated sieve runs (since the run-dependent component of variability is not considered), they provide a useful means of comparing the repeatability of the inherently quantised sieve measurements with shape features derived from silhouette sets on an individual measurement basis.

The ΣVAR values indicate that visual hull volume is more repeatable than sieving in terms of individual measurements. Since visual hull repeatability is not run-dependent, while sieve repeatability is run-dependent, visual hull volume histograms are also expected to be more repeatable than sieving histograms.

The table confirms observations of measurement repeatability implied by matching accuracy (see Table 10.2): (1) VEMH-based measurements tend to be more repeatable than visual hull-based measurements, (2) volume measurements tend to be more repeatable than minimum cylinder measurements.

Only the minimum enclosing cylinder of the visual hull appears to be less repeatable than the sieve-based measurements on an individual stone basis. This suggests that the minimum enclosing cylinder of the VEMH (rather than the visual hull) should be used to emulate sieving.

The contribution that each stone makes to the ΣVAR statistic can be used as a measure of how likely the stone is to land in different bins on different runs. Presumably, the shape of a stone plays an important role in determining the likelihood that a stone will have a tendency to fall into different bins on different runs. Figure 10.7 illustrates the five stones that produced the largest contribution to the sieving ΣVAR statistic. The 3D shapes do not seem to provide any obvious clues as to why these stones have a tendency to fall into

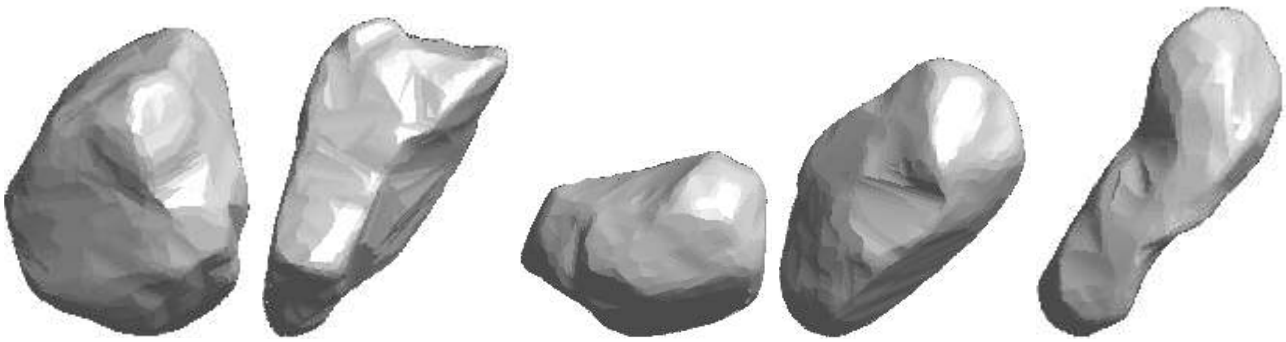


Figure 10.7: Refined visual hull models of stones that tend to fall in different bins as measured by contribution to ΣVAR.

different bins. However, the stones do exhibit some protrusions which may cause the stone to become stuck in an aperture in certain orientations.

Also note that a stone whose minimum sieve aperture is just larger than an actual sieve aperture will have less of a tendency of fall into different bins on different runs (as it will easily pass through the bin's upper boundary but cannot pass through the lower boundary). The actual bin boundaries also therefore play a role in determining the ΣVAR contribution for a stone.

10.5 Summary

An experiment in which 494 garnets were sieved 15 times has been presented. The experiment makes use of the main shape, calibration, and recognition methods developed in this thesis to compare sieve sizing with silhouette-based estimates of shape properties.

The proposed batch matching method (see Chapter 9) has been demonstrated to produce perfectly consistent matches over all run pairs. This is in contrast to other simpler methods that all exhibit inconsistency, thereby providing justification for the additional complexity of the proposed batch matching method. Matching a pair of runs of 494 stones takes approximately three seconds (in addition to the preprocessing that can be carried out as the stones are passed through the system). The running time is therefore, for practical purposes, insignificant.

The minimum enclosing cylinder estimator has been used to emulate sieve measurements. Histograms of sieve measurements were compared with those of silhouette-based sieve emulators. Refined visual hull models formed from 90 views (15 runs of six-view sets) demonstrated the limitations of using minimum cylinders to predict sieve bins. (Note that batch matching provides an efficient means of obtaining the 90-view visual hulls: without batch matching, each stone would have to be individually passed through the camera setup 15 times; batch matching allows the stones to be passed through in batches, substantially speeding up data capture.) Cases were found in which the minimum cylinder diameter of a 90-view visual hull was up to 10% larger than the sieve aperture of the stone's minimum bin. However, the minimum cylinder diameter was larger than the minimum bin's lower boundary diameter in all cases.

The sum of bin count variances (ΣVAR) has been introduced as a means for comparing the repeatability of silhouette-based shape properties with sieving, which produces histograms as output. Volume estimates based on visual hull volume, and sieve emulation based on the minimum enclosing cylinder of the VEMH have been shown to be more repeatable than sieving for the data captured using a batch of 494 garnets.

Chapter 11

Conclusion

11.1 Summary of Contributions

This thesis has extended the capabilities of multi-view silhouette based particle analysis by incorporating silhouette consistency constraints. Three problems have been addressed: (1) camera calibration, (2) estimating shape, and (3) recognising individual stones.

To provide practical tools to particle shape analysts, running time efficiency has been considered: computations which require hours or days to complete are impractical. ET error and the VEMH have played an important role in the design of efficient methods throughout this thesis. ET error is an efficiently computable measure of silhouette consistency, and a VEMH is an efficiently computable estimate of the shape that produced a silhouette set.

11.1.1 Calibration

The configuration and calibration of two image capture setups have been addressed. The first, the two-mirror setup, is a low cost setup that can be easily created using readily available equipment. The second, the six-camera setup, is a high throughput system that can be used for large batches of stones in either an industrial or a laboratory setting.

The Two-Mirror Setup

The two-mirror setup is used to capture five silhouette views of an object in a single image. It has been shown that the setup can be calibrated using only constraints imposed by silhouette bitangents. This approach therefore adds to the array of silhouette-based self-calibration methods described in the computer vision

literature. These approaches include setups that impose constraints based on known circular motion [93] and prior knowledge of camera orientation [102].

Calibration involves determining the camera pose and internal parameters as well as mirror poses; there is therefore no need for accurate positioning of any apparatus. The two-mirror setup provides a convenient approach to capturing multiple calibrated silhouette views of stones for shape-from-silhouette reconstruction without using specialised equipment. Its use is not, however, limited to stones: it can be used for shape reconstruction of arbitrary objects. Experiments have demonstrated that calibration is sufficiently accurate that silhouette noise is a greater contribution to inconsistency across silhouettes than calibration parameter errors.

The Six-Camera Setup

Two different heuristics were considered for determining the camera configuration for the six-camera setup: one requires maximising the distribution of frontier points on a sphere, and the other minimises the isolation of the direction that is furthest from any viewing direction. Both heuristics are designed to provide good results over a range of silhouette-based applications (estimating shape, volume, and matching), and both indicate that six cameras should be configured so that viewing directions are perpendicular to the parallel face pairs of a regular dodecahedron. This is therefore the configuration that is used.

The six-camera setup is calibrated using several runs of silhouette sets of a ball. Initial parameter estimates are computed by generating approximate point correspondences using the centres of the ball projections. The method is based on the work of Tomasi and Kanade [129]. The calibration parameters are then refined by minimising ET error, and scale is enforced using the known size of the ball.

Merging Silhouette Sets

A method for aligning silhouette sets in a common reference frame by minimising ET error has been introduced. This is external calibration: the poses of the cameras must be specified, but the internal parameters are known. The method allows silhouette sets containing a large number of views of a stone to be constructed from setups that produce a small number of views (such as the two-mirror setup and the six-camera setup). A larger number of silhouettes provides more constraints on stone shape, which provides the potential for more accurate estimates of shape properties.

11.1.2 Recognition

The main recognition goal of this thesis is efficient batch matching: an algorithm to compute the one to one correspondences between two unordered batches of silhouette sets of the same batch of stones. Batch

matching is useful for tasks such as reconciling class labels assigned to each stone using batch classification methods such as sieving (in which stones are classified together rather than individually). The design of a batch matching algorithm was split into three components: (1) alignment-based matching, (2) faster, but less accurate signature-based matching, and (3) a framework to combine the accuracy of alignment-based matching with the speed of signature-based matching to create an efficient batch matching algorithm.

Alignment-Based Matching

Alignment-based matching simply applies ET-based pose optimisation to a pair of silhouettes. If a sufficiently low error is achieved, then the pair is classified as a match, otherwise it is classified as a mismatch. A formulation of ET error based on an orthographic projection model was introduced to improve computational efficiency. The method was applied to the 2-mirror 5-view gravel data set and the 6-camera garnet data set. All mismatch pairs considered were found to produce substantially larger ET errors after alignment than any of the match pairs considered. Various CIP-based error formulations were found to produce no practical improvement on matching accuracy when tested on downsampled image data. Used alone, alignment-based matching provides a means for verification: a silhouette set of a stone can be compared with a silhouette set on record to confirm that the two silhouette sets correspond to the same stone.

Signature-Based Matching

Signature-based matching uses signatures that approximate the CDF of a stone's caliper diameter distribution. The EMD between signature pairs is used to quantify their dissimilarity. The EMD between signatures can be computed in $O(m)$ time complexity for m-element signatures (typically $m = 10$), and in practice takes less than one microsecond to compute. This makes it practical to compute dissimilarity values between all n^2 pairings across two runs of n stones (for realistic values of n; a batch of stones will contain several thousand stones at the most). Although signature-based matching was developed primarily as a component of batch matching, it can also be used for identification. To identify a query silhouette set from a database of stored silhouette sets, the query-database pairings can be rapidly ranked in order of dissimilarity specified by EMD between signatures. The slower alignment-based matching is then applied to pairs in order of dissimilarity. Tests applied to the 2-mirror 5-view gravel data set result in the correct match being ranked first by EMD in 98% of cases, and the correct match is always within the top five.

Batch Matching

A simple probabilistic framework was used for batch matching. Each silhouette set pair across two runs is assigned a likelihood ratio (indicating the likelihood of being a match). The pairs are pushed onto a priority

queue that is prioritised by likelihood ratio. Alignment-based matching is used to make a hard (i.e., non-probabilistic) decision for each pair and starting point considered: match or unknown. Once all matches are found, the algorithm terminates. Efficiency is achieved by reducing the likelihood ratio using Bayes's rule and pushing the pair back into the priority queue after an unsuccessful optimisation. This is a greedy algorithm: pose optimisation is always applied to the pair and starting point that is most likely to result in successful alignment. A data set of two runs of six-view silhouette sets of uncut gemstones is correctly matched up in approximately 68 seconds on a 3.2 GHz Pentium 4 machine. Of this, 50 seconds is spent on preprocessing that can be computed online as the stones pass through the six-camera setup.

11.1.3 Shape

The VEMH has been introduced as an alternative to the visual hull for estimating the shape of the convex hull of a stone from its silhouettes. The VEMH can be used to estimate the caliper diameter of a stone in a given direction. This has been used for recognising stones, but is also of use to particle shape analysts who use estimates of the short, intermediate, and long diameters for a broad range of applications.

The accuracy with which commonly-used shape properties (long, intermediate, short diameters and volume) can be estimated from silhouette sets has been quantified for both image capture setups considered. Merging silhouette sets to create a single large silhouette set of a stone from silhouette sets containing a smaller number of silhouettes has been shown to improve the accuracy in estimating these shape properties.

The extent to which the minimum enclosing cylinder can be used to emulate sieving has been investigated in an experiment which makes use of the calibration, recognition, and shape methods developed in this thesis. The sieve bin associated with each of 494 garnets across 15 runs of sieving was determined using batch matching. The bins associated with most stones are consistent with the minimum cylinder diameter, limiting the smallest sieve aperture through which the stone can pass. A few stones landed in bins bounded by circular apertures with diameters smaller than the minimum cylinder, indicating that these stones may have changed their direction of motion as they passed through the aperture. Both visual hull-based volume estimates and silhouette-based sieve emulation were found to produce more repeatable histograms than sieving for the data set of 494 garnets.

11.2 Future Work

There are many ways in which the work described in this thesis can be extended. Some ideas follow.

This work has been limited to considering silhouette images of stones. Front-lit images from colour cameras will provide information about the colour and surface texture of stones. This information may enable estimation of particle properties that are not available from silhouettes. Stereo vision techniques may be able to reconstruct concavities that cannot be captured with silhouettes.

There has recently been interest in the discrete element modelling community in using silhouette-based methods for model validation [107]. Li et al. [79] use computer simulations with discrete element modelling. They are interesting in the efficiency of the sieving process as a function of sieving time and intensity. The methods presented in this thesis may be useful for validating this type of computer simulation with real experiments. For instance, the results of a computer simulation of sieving (using 3D shape models computed from a real batch of stones) can be compared with the results that are achieved in practice on a stone by stone basis.

Some initial work indicates that ET-based alignment of silhouette sets may be adapted to align silhouettes of a stone before and after the stone is chipped. Figure 11.1 shows an example.

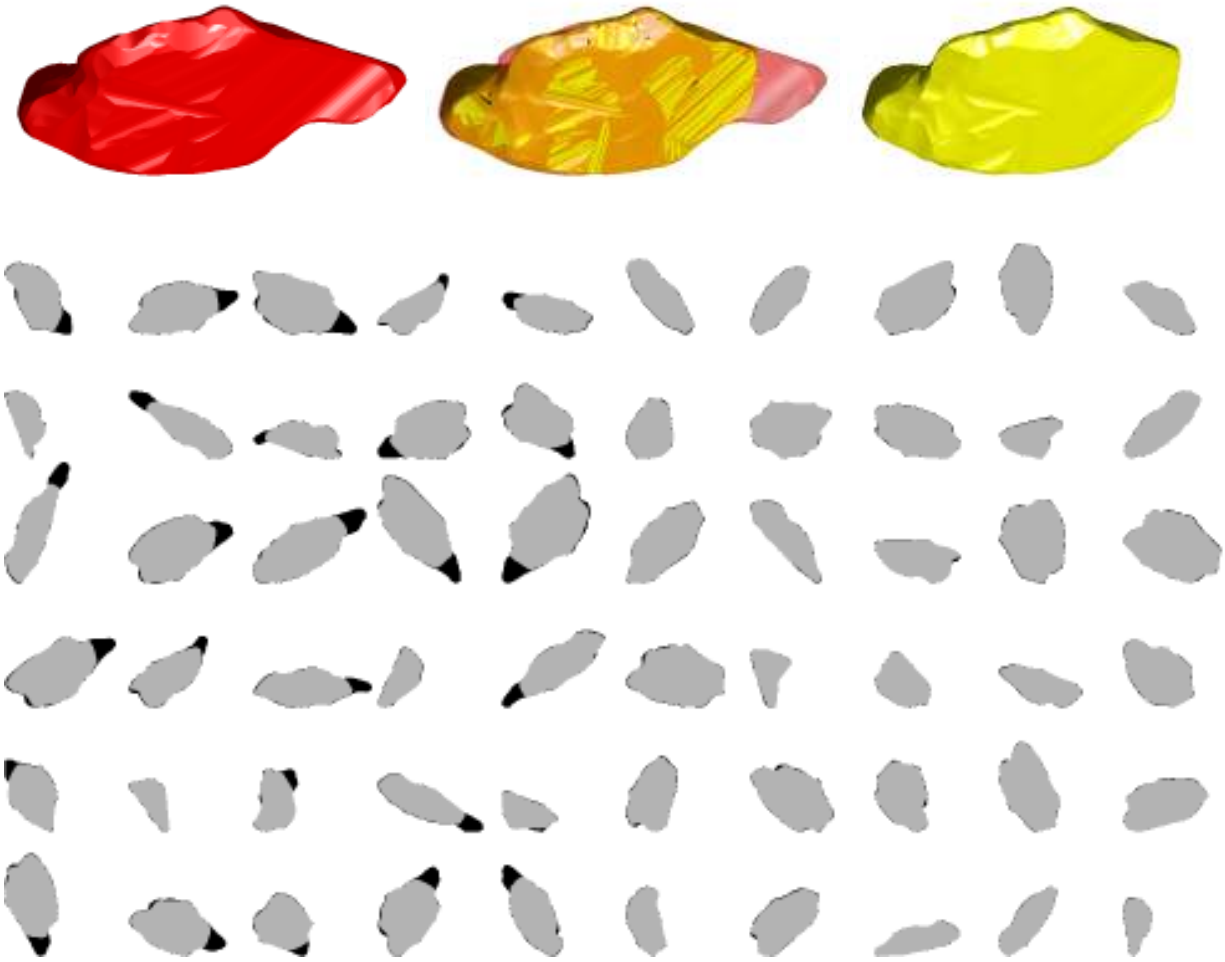


Figure 11.1: A refined visual hull model of a stone formed by merging five 6-view runs is shown in red. A portion of the stone was subsequently chipped off. A refined visual hull model of the chipped stone formed by merging five 6-view runs is shown in yellow. The original stone is also overlaid on the broken version to aid visualisation of the chipped piece. Silhouettes of the stone are shown in black with the CIPs formed from all silhouettes overlaid in grey. Each row of silhouettes corresponds to a camera view, and each column, a run. The stone was chipped between runs 5 and 6. CIPs therefore do not cover portions of the original stone corresponding to the chipped piece: these portions are black.

To specify all silhouettes in a common reference frame, pose optimisation using a modified version of ET error that accounts for the chipping was used. (The modification assumes stone projections can become smaller, but not larger between runs 5 and 6.) This provides the potential to analyse the shape and location of chipping during certain industrial processes, and to recognise stones even if they are chipped. Understanding the nature of breakage is important when dealing with high value gemstones. By identifying the shape and location of chips broken from real stones, it may be possible to validate computer simulations that attempt to predict the breakage occurrences. Since constraints on relative pose are weaker after breakage, it may be necessary to use more silhouettes or to incorporate CIP constraints in addition to ET constraints.

Gemstones are manually classified into different shape classes for valuation purposes. It is possible that some of the recognition methods developed in this thesis could be extended to automate the shape classification of gemstones.

Appendix A

Threshold-Based Subpixel Segmentation

The methods developed in this thesis use as input polygonal boundaries that separate the foreground and background regions in images. This appendix describes the silhouette boundary extraction algorithm used to segment the images captured by the six-camera setup described in Chapter 5. Since images are captured under controlled lighting conditions, a simple threshold-based segmentation effectively separates foreground from background.

An empty background image is stored for each of the six cameras so that background subtraction can be applied. This reduces the effect of any intensity variation of the background over an image. Otsu's method [105] is applied to the difference images to determine thresholds. The method selects a threshold to minimise the intra-class variance of pixel intensity values for background and foreground. In practice, the extracted boundaries are found to be insensitive to the precise threshold value since backlights ensure that background pixels are substantially brighter than foreground pixels.

The algorithm achieves efficiency by using a strategy that does not visit each pixel. This can be done because the boundary of only one connected region is sought per image (i.e., the prior knowledge that each image contains exactly one stone silhouette is used). Only visited pixels are classified as foreground or background. This is done by subtracting the pixel intensity value of the background image from the pixel intensity value of the foreground image and comparing the result with the fixed threshold value. This allows segmentation to be carried out without visiting each pixel. After a pixel-resolution boundary is extracted, the boundary is traversed once more to compute a subpixel resolution boundary using linear interpolation of pixel intensity values. The resultant boundary is equivalent to the *marching squares* boundary. (Marching squares is the 2D analogue of marching cubes [84].)

Broadly, the algorithm proceeds as follows:

1. Find a pixel *inside* the silhouette.
2. Walk downwards to find the boundary of the silhouette.

3. Traverse the silhouette to determine its pixel resolution boundary.

4. Traverse the silhouette again to determine its subpixel resolution boundary.

A.1 Finding a Starting Point

The first step of the algorithm is to find a pixel that is sufficiently dark (i.e., a foreground pixel). This is carried out by considering grid point vertices at successively finer resolutions until a foreground pixel is found. The first point is the image centre. This point belongs to the Level 1 Grid. The Level 2 Grid is formed by points in the centres of the four rectangles defined by the image corners and the first vertex. There are therefore four Level 2 vertices. Vertices belonging to further grid levels are defined in a similar manner using the rectangle centres of the grid's predecessor. Figure A.1 illustrates the point locations for grids up to Level 5. There are $2^{2(n-1)}$ point locations for a grid of level n.

The procedure of looking for a foreground pixel ensures that few pixels are visited. Figure A.2 shows an example. In this case, ten grid points are considered before a sufficiently dark pixel is found.

Once a dark pixel has been found, the algorithm searches for a background pixel. This is done by moving downwards one pixel at a time as can be seen in the example in Figure A.3.

A.2 Pixel-Resolution Boundary

Once a pair of foreground and background pixels has been found, their shared edge is used as the first edge of the boundary polygon, and the leftmost vertex of this edge is used as the first vertex of the polygon (see Figure A.3).

The polygon is then traversed by moving from pixel corner to pixel corner, keeping the silhouette to the left. At each step the boundary can proceed either left, straight ahead, or right. This is determined by considering the two pixels ahead of the current polygon edge (the *ahead left pixel* and the *ahead right pixel*). The rules for determining the direction of the next edge from the current edge vertex are given in Table A.1. The rules imply that the foreground is 4-connected and the background is 8-connected. However, in practice, there are rarely images for which a 4-connected foreground differs from an 8-connected foreground.

ahead left pixel	ahead right pixel	direction
foreground	foreground	right
foreground	background	straight ahead
background	foreground	left
background	background	left

Table A.1: Determining the direction of the next edge from the *ahead left pixel* and the *ahead right pixel*.

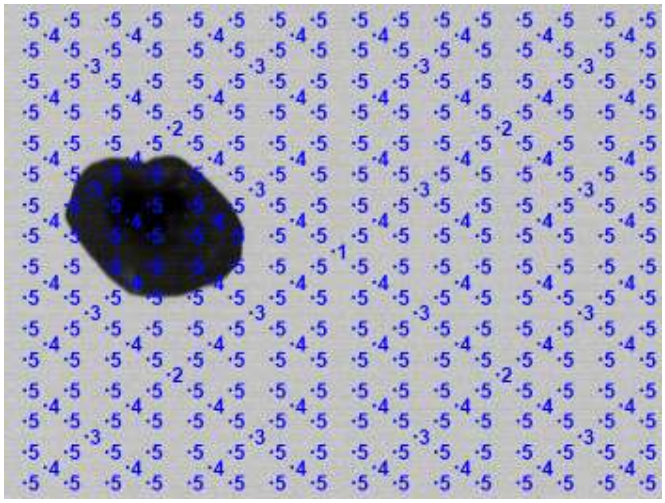


Figure A.1: Example showing point locations considered when searching for a foreground pixel. The grid level number associated with each point is shown next to each point.

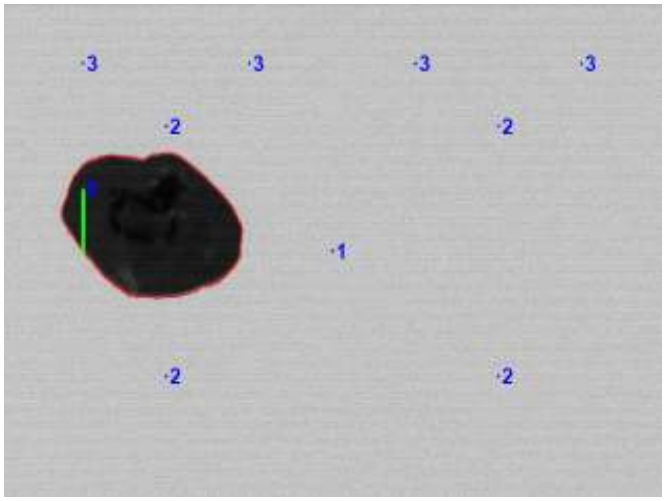


Figure A.2: A segmentation example. Grid points used to locate a foreground pixel are in blue with the level number shown. The vertical path from the starting point to the boundary is in green, and the silhouette boundary is red.

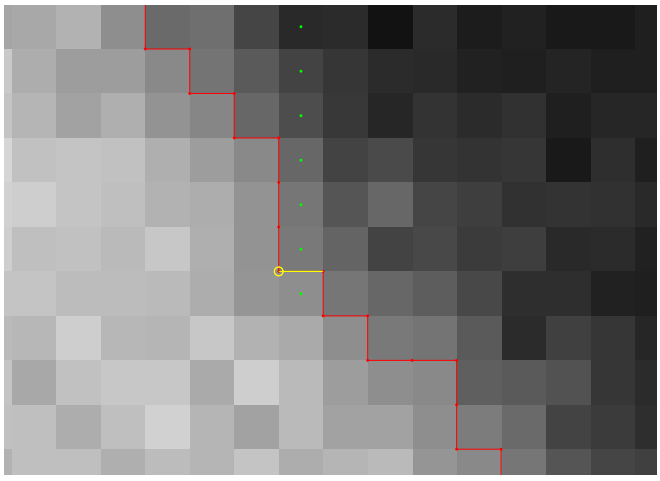


Figure A.3: A closeup of part of Figure A.2. The vertical path from the foreground starting point to the boundary is shown with green dots. The boundary is in red. A yellow circle indicates the start of the boundary, and a yellow line segment indicates the first edge of the polygon representing the boundary.

Upon return to the starting point, the pixel resolution boundary has been found.

A.3 Subpixel Boundary

Figure A.4 shows an example of a portion of a subpixel boundary. The original boundary (shown in green) runs across pixel edges and is therefore limited to pixel resolution. To create a subpixel boundary, linear interpolation based on pixel intensity values is used. For each edge, a pixel-length line segment is considered. Each line segment runs from the centre of one of the pixels bounded by the edge to the centre of the other pixel bounded by the edge. One pixel is a background pixel, and one is a foreground pixel. A vertex of the subpixel boundary is generated along each line segment. The vertex is positioned at a distance p from the centre of the foreground pixel using the following formula:

$$p = \frac{i_F - i_T}{i_F - i_B}, \tag{A.1}$$

where i_F is the intensity value of the foreground pixel, i_B is the intensity value of the background pixel, and i_T is the threshold value.

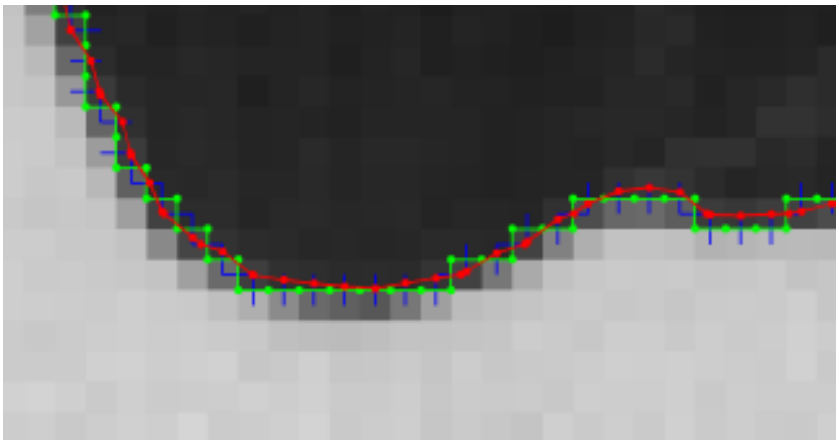
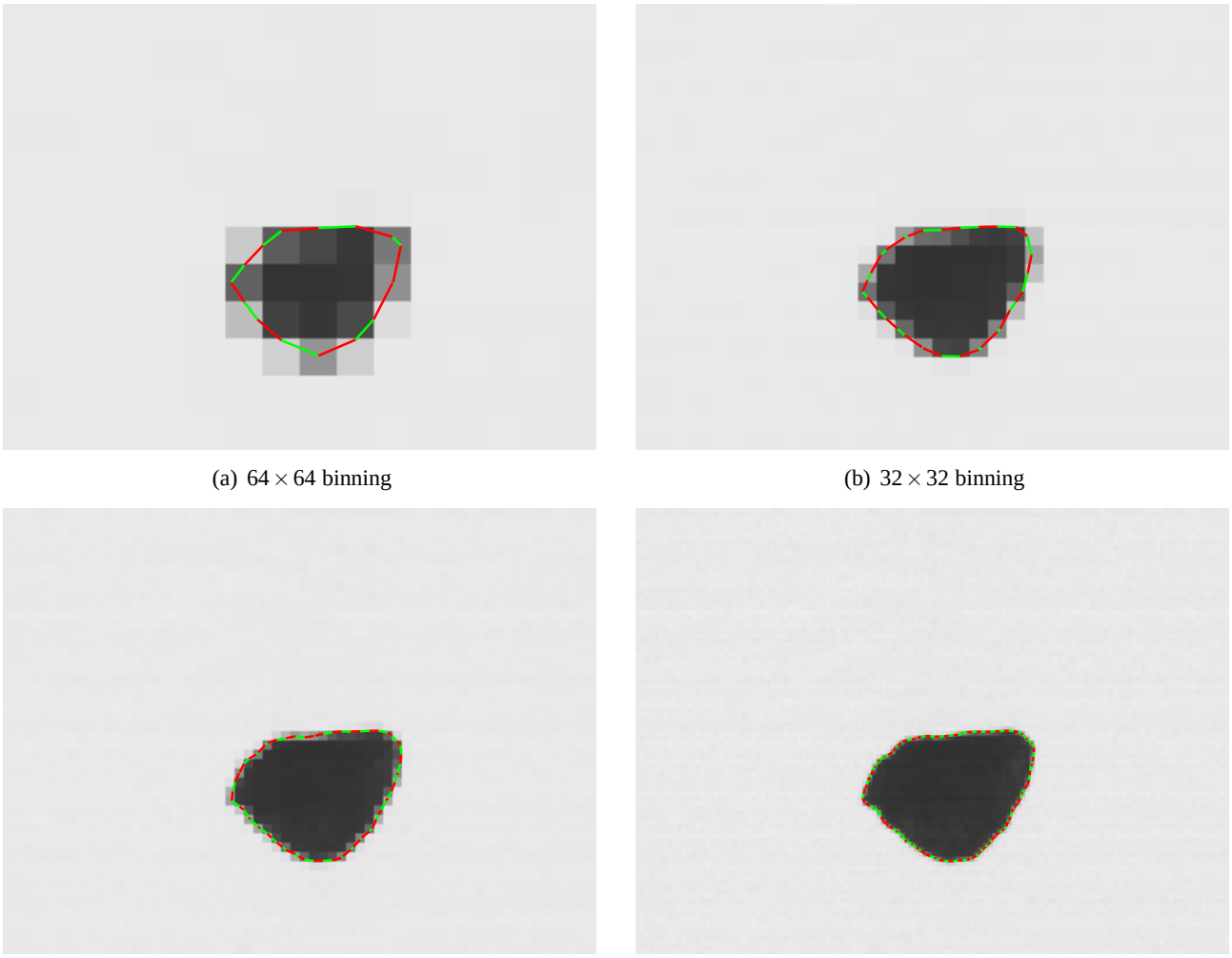


Figure A.4: An example of subpixel segmentation. The original pixel resolution boundary is shown in green. The subpixel boundary is shown in red. Each vertex of the subpixel boundary lies on a pixel-length line segment associated with each edge of the original boundary. These line segments (shown in blue) are perpendicular to and share centre points with the associated original edges.

The resultant subpixel boundary is the same as a marching squares boundary. The marching squares algorithm considers pixel-sized squares centred at each pixel corner. The corners of the squares (which lie on pixel centres) are classified as foreground or background based on the intensity threshold i_T. The $2^4 = 16$ possible classifications of the square determine how the boundary enters and exits the square, i.e., which square sides the boundary crosses. The square sides are equivalent to the pixel-length line segments considered above.

A.4 Experiments

To test the performance of subpixel segmentation, image downsampling was used. This was done because intensity noise in the images is high, and at full resolution intensity noise dominates spatial quantisation noise. Downsampling uses the mean intensity value of $n \times n$ groups of pixels to create a lower resolution image. Averaging the intensity values has the effect of reducing intensity noise, but increasing spatial quantisation noise.



(a) 64×64 binning

(b) 32×32 binning

(c) 16×16 binning

(d) 8×8 binning

Figure A.5: Subpixel segmentation of an image of a garnet with various levels of downsampling. The subpixel boundary is shown in colour.

Figure A.5 shows examples of $n \times n$ pixel binning of a garnet image for various values of n.

Pixel binning was applied to 246 image sets of garnets captured with the six-camera setup. Bin dimensions of 1×1 (i.e., no binning) through to 20×20 were used. Boundaries were then extracted using both pixel resolution and subpixel boundary extraction methods (with fixed intensity thresholds determined by the

calibration procedure). At full resolution, the C++ implementation of the subpixel boundary extraction takes approximately 1.5 ms per image on a 3.2 GHz Pentium 4 machine.

After extraction, the boundaries were scaled up by a factor of n to facilitate direct comparison between different degrees of downsampling, and so that unaltered computed calibration parameters could be used. Figure A.6 shows a plot of the internal ET error for the silhouette sets at different resolution levels using pixel resolution boundaries and using subpixel boundaries. The plot shows that with no downsampling

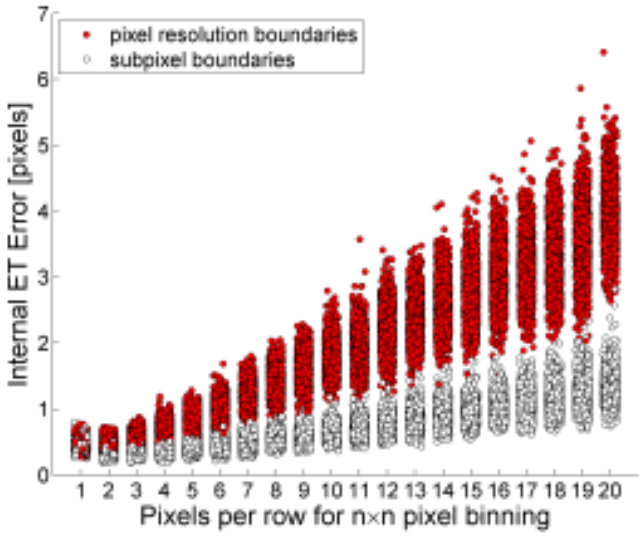


Figure A.6: Plot of internal ET error versus n for $n \times n$ pixel binning of 246 silhouette sets of garnets.

(1×1 binning), subpixel boundary extraction offers no greater accuracy than pixel resolution boundary extraction. This is because intensity noise dominates spatial quantisation noise. With 2×2 binning the ET error *decreases*. This is because the averaging effect that reduces pixel intensity noise outweighs the increased spatial quantisation noise. This suggests that, given the current lighting, it would be beneficial to run the cameras of the current six-camera setup in 2×2 binning mode (the Dragonfly cameras used with the current setup can be configured to run in this mode). This also indicates that there is scope for improvement in boundary accuracy if the signal to noise ratio of the images is improved, by using brighter backlights for instance.

The plot illustrates that subpixel boundary extraction produces significantly more accurate boundaries (in terms of ET error) than pixel resolution boundaries when spatial quantisation errors are the major source of error.

Appendix B

An Analytical Expression for a Jacobian Matrix

This appendix records the derivation of an analytical expression for the Jacobian matrix that is used for ET-based pose optimisation with an orthographic imaging model (as described in Chapter 7).

The Jacobian matrix is used by the Levenberg-Marquardt method to create a local linear approximation to the cost function

$$\mathbf{f}(\mathbf{v}) = \mathbf{e}. \tag{B.1}$$

In this case, $\mathbf{v} = (q_1, q_2, q_3, q_4, x_t, y_t, z_t)^T$ is a seven parameter vector representing a pose. The orientation part of the pose is represented by a (possibly non-unit) quaternion (q_1, q_2, q_3, q_4). To determine the rotation, the quaternion is unitised. The *x*-, *y*-, and *z*-components of the translational part of the pose is represented by x_t, y_t and z_t. The error vector **e** stores the individual residual values:

$$\mathbf{e} = (\Delta x_{0A1B1}, \Delta y_{0A1B1}, \Delta x_{1A1B1}, \Delta y_{1A1B1} \ldots \Delta x_{0AmBn}, \Delta y_{0AmBn}, \Delta x_{1AmBn}, \Delta y_{1AmBn})^T. \tag{B.2}$$

Each residual value is identified by subscripts. The first indicates to which of the two outer tangencies the reprojection error corresponds (0 or 1). Subsequent subscripts indicate which image from Set A and which image from Set B correspond to the residual value.

The Jacobian matrix J is as follows:

$$\mathsf{J} = \begin{pmatrix} \frac{\partial \Delta x_{0A1B1}}{\partial q_1} & \frac{\partial \Delta x_{0A1B1}}{\partial q_2} & \frac{\partial \Delta x_{0A1B1}}{\partial q_3} & \frac{\partial \Delta x_{0A1B1}}{\partial q_4} & \frac{\partial \Delta x_{0A1B1}}{\partial x_t} & \frac{\partial \Delta x_{0A1B1}}{\partial y_t} & \frac{\partial \Delta x_{0A1B1}}{\partial z_t} \\ \frac{\partial \Delta y_{0A1B1}}{\partial q_1} & \frac{\partial \Delta y_{0A1B1}}{\partial q_2} & \frac{\partial \Delta y_{0A1B1}}{\partial q_3} & \frac{\partial \Delta y_{0A1B1}}{\partial q_4} & \frac{\partial \Delta y_{0A1B1}}{\partial x_t} & \frac{\partial \Delta y_{0A1B1}}{\partial y_t} & \frac{\partial \Delta y_{0A1B1}}{\partial z_t} \\ \frac{\partial \Delta x_{1A1B1}}{\partial q_1} & \frac{\partial \Delta x_{1A1B1}}{\partial q_2} & \frac{\partial \Delta x_{1A1B1}}{\partial q_3} & \frac{\partial \Delta x_{1A1B1}}{\partial q_4} & \frac{\partial \Delta x_{1A1B1}}{\partial x_t} & \frac{\partial \Delta x_{1A1B1}}{\partial y_t} & \frac{\partial \Delta x_{1A1B1}}{\partial z_t} \\ \frac{\partial \Delta y_{1A1B1}}{\partial q_1} & \frac{\partial \Delta y_{1A1B1}}{\partial q_2} & \frac{\partial \Delta y_{1A1B1}}{\partial q_3} & \frac{\partial \Delta y_{1A1B1}}{\partial q_4} & \frac{\partial \Delta y_{1A1B1}}{\partial x_t} & \frac{\partial \Delta y_{1A1B1}}{\partial y_t} & \frac{\partial \Delta y_{1A1B1}}{\partial z_t} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \frac{\partial \Delta x_{0AmBn}}{\partial q_1} & \frac{\partial \Delta x_{0AmBn}}{\partial q_2} & \frac{\partial \Delta x_{0AmBn}}{\partial q_3} & \frac{\partial \Delta x_{0AmBn}}{\partial q_4} & \frac{\partial \Delta x_{0AmBn}}{\partial x_t} & \frac{\partial \Delta x_{0AmBn}}{\partial y_t} & \frac{\partial \Delta x_{0AmBn}}{\partial z_t} \\ \frac{\partial \Delta y_{0AmBn}}{\partial q_1} & \frac{\partial \Delta y_{0AmBn}}{\partial q_2} & \frac{\partial \Delta y_{0AmBn}}{\partial q_3} & \frac{\partial \Delta y_{0AmBn}}{\partial q_4} & \frac{\partial \Delta y_{0AmBn}}{\partial x_t} & \frac{\partial \Delta y_{0AmBn}}{\partial y_t} & \frac{\partial \Delta y_{0AmBn}}{\partial z_t} \\ \frac{\partial \Delta x_{1AmBn}}{\partial q_1} & \frac{\partial \Delta x_{1AmBn}}{\partial q_2} & \frac{\partial \Delta x_{1AmBn}}{\partial q_3} & \frac{\partial \Delta x_{1AmBn}}{\partial q_4} & \frac{\partial \Delta x_{1AmBn}}{\partial x_t} & \frac{\partial \Delta x_{1AmBn}}{\partial y_t} & \frac{\partial \Delta x_{1AmBn}}{\partial z_t} \\ \frac{\partial \Delta y_{1AmBn}}{\partial q_1} & \frac{\partial \Delta y_{1AmBn}}{\partial q_2} & \frac{\partial \Delta y_{1AmBn}}{\partial q_3} & \frac{\partial \Delta y_{1AmBn}}{\partial q_4} & \frac{\partial \Delta y_{1AmBn}}{\partial x_t} & \frac{\partial \Delta y_{1AmBn}}{\partial y_t} & \frac{\partial \Delta y_{1AmBn}}{\partial z_t} \end{pmatrix} \tag{B.3}$$

The Jacobian matrix J has as many rows as there are elements in **e** and as many columns as there are elements in **v**. Since there are reprojection errors in the *x*- and *y*-directions for each of two outer tangencies, there are $2 \times 2 = 4$ reprojection errors for each silhouette pair. Note that the residual values are computed in only one image of the pair. Since we are using an orthographic projection, the corresponding residuals computed in the opposite image are identical.

If Set A contains *m* silhouettes and Set B contains *n* silhouettes, then there are *mn* pairings across sets. There are thus 4*mn* reprojection errors corresponding to a pose estimate.

The partial derivatives that are the elements of J are computed using the chain rule. For example

$$\frac{\partial \Delta x}{\partial q_1} = \frac{\partial \Delta x}{\partial r_{11}}\frac{\partial r_{11}}{\partial q_1} + \frac{\partial \Delta x}{\partial r_{12}}\frac{\partial r_{12}}{\partial q_1} + \frac{\partial \Delta x}{\partial r_{13}}\frac{\partial r_{13}}{\partial q_1} + \frac{\partial \Delta x}{\partial t_x}\frac{\partial t_x}{\partial q_1} + \frac{\partial \Delta x}{\partial r_{21}}\frac{\partial r_{21}}{\partial q_1} + \frac{\partial \Delta x}{\partial r_{22}}\frac{\partial r_{22}}{\partial q_1} + \frac{\partial \Delta x}{\partial r_{23}}\frac{\partial r_{23}}{\partial q_1} + \frac{\partial \Delta x}{\partial t_y}\frac{\partial t_y}{\partial q_1}. \tag{B.4}$$

The Matlab Symbolic Math Toolbox was used as an aid in computing the following partial derivatives:

$$\frac{\partial \Delta x}{\partial r_{11}} = \frac{{r_{23}}^2 p_{2x}}{{r_{23}}^2 + {r_{13}}^2} \tag{B.5}$$

$$\frac{\partial \Delta x}{\partial r_{12}} = \frac{{r_{23}}^2 p_{2y}}{{r_{23}}^2 + {r_{13}}^2} \tag{B.6}$$

$$\frac{\partial \Delta x}{\partial t_x} = \frac{{r_{23}}^2}{{r_{23}}^2 + {r_{13}}^2} \tag{B.7}$$

$$\frac{\partial \Delta x}{\partial r_{13}} = \frac{\left(p_{1y} - r_{21}\, p_{2x} - r_{22}\, p_{2y} - t_y\right) r_{23}}{{r_{23}}^2 + {r_{13}}^2} + 2\,\frac{r_{23}^2 r_{13}\left(p_{1x} - r_{11}\, p_{2x} - r_{12}\, p_{2y} - t_x\right) - r_{13}^2 r_{23}\left(p_{1y} - r_{21}\, p_{2x} - r_{22}\, p_{2y} - t_y\right)}{\left({r_{23}}^2 + {r_{13}}^2\right)^2} \tag{B.8}$$

$$\frac{\partial \Delta x}{\partial r_{21}} = -\frac{r_{13}\, p_{2x}\, r_{23}}{{r_{23}}^2 + {r_{13}}^2} \tag{B.9}$$

$$\frac{\partial \Delta x}{\partial r_{22}} = \frac{r_{13}\, p_{2y}\, r_{23}}{{r_{23}}^2 + {r_{13}}^2} \tag{B.10}$$

$$\frac{\partial \Delta x}{\partial t_y} = -\frac{r_{13}\, r_{23}}{{r_{23}}^2 + {r_{13}}^2} \tag{B.11}$$

$$\frac{\partial \Delta x}{\partial r_{23}} = \frac{r_{13}\left(p_{1y} - r_{21}\,p_{2x} - r_{22}\,p_{2y} - t_y\right) - 2r_{23}\left(p_{1x} - r_{11}\,p_{2x} - r_{12}\,p_{2y} - t_x\right)}{{r_{23}}^2 + {r_{13}}^2} + \frac{r_{23}^3\left(p_{1x} - r_{11}\,p_{2x} - r_{12}\,p_{2y} - t_x\right) - r_{13}r_{23}^2\left(p_{1y} - r_{21}\,p_{2x} - r_{22}\,p_{2y} - t_y\right)}{\left({r_{23}}^2 + {r_{13}}^2\right)^2} \quad \text{(B.12)}$$

$$\frac{\partial \Delta y}{\partial r_{11}} = -\frac{r_{23}\,p_{2x}\,r_{13}}{{r_{23}}^2 + {r_{13}}^2} \quad \text{(B.13)} \qquad \frac{\partial \Delta y}{\partial r_{12}} = -\frac{r_{23}\,p_{2y}\,r_{13}}{{r_{23}}^2 + {r_{13}}^2} \quad \text{(B.14)} \qquad \frac{\partial \Delta y}{\partial t_x} = -\frac{r_{23}\,r_{13}}{{r_{23}}^2 + {r_{13}}^2} \quad \text{(B.15)}$$

$$\frac{\partial \Delta y}{\partial r_{13}} = \frac{\left(r_{23}^3 - r_{13}^2 r_{23}\right)\left(p_{1x} - r_{11}\,p_{2x} - r_{12}\,p_{2y} - t_x\right) - 2r_{13}r_{23}^2\left(p_{1y} - r_{21}\,p_{2x} - r_{22}\,p_{2y} - t_y\right)}{\left({r_{23}}^2 + {r_{13}}^2\right)^2} \quad \text{(B.16)}$$

$$\frac{\partial \Delta y}{\partial r_{21}} = \frac{{r_{13}}^2 p_{2x}}{{r_{23}}^2 + {r_{13}}^2} \quad \text{(B.17)} \qquad \frac{\partial \Delta y}{\partial r_{22}} = \frac{{r_{13}}^2 p_{2y}}{{r_{23}}^2 + {r_{13}}^2} \quad \text{(B.18)} \qquad \frac{\partial \Delta y}{\partial t_y} = \frac{{r_{13}}^2}{{r_{23}}^2 + {r_{13}}^2} \quad \text{(B.19)}$$

$$\frac{\partial \Delta y}{\partial r_{23}} = \frac{r_{13}\left(p_{1x} - r_{11}\,p_{2x} - r_{12}\,p_{2y} - t_x\right)}{{r_{23}}^2 + {r_{13}}^2} - 2\,\frac{r_{23}^2 r_{13}\left(p_{1x} - r_{11}\,p_{2x} - r_{12}\,p_{2y} - t_x\right) - r_{13}^2 r_{23}\left(p_{1y} - r_{21}\,p_{2x} - r_{22}\,p_{2y} - t_y\right)}{\left({r_{23}}^2 + {r_{13}}^2\right)^2}. \quad \text{(B.20)}$$

A candidate rigid transform from View 2 of Set B to View 1 of Set A is derived from its parameter vector $\mathbf{v}$ and the poses of Views 1 and 2 with respect to the world reference frames of Sets A and B.

A 4×4 matrix representing a rigid transform is derived from the parameter vector $\mathbf{v}$. The first four elements of $\mathbf{v}$ represent a quaternion. The normalised (unit) quaternion $\hat{\mathbf{q}} = (q_x, q_y, q_z, q_w)^T$ represents a rotation:

$$\begin{pmatrix} q_x \\ q_y \\ q_z \\ q_w \end{pmatrix} = \frac{1}{\sqrt{q_1^2 + q_2^2 + q_3^2 + q_4^2}} \begin{pmatrix} q_1 \\ q_2 \\ q_3 \\ q_4 \end{pmatrix}. \quad \text{(B.21)}$$

A rotation matrix $\mathsf{R}_{\mathrm{B}\to\mathrm{A}}$ is formed from the unit quaternion:

$$\mathsf{R}_{\mathrm{B}\to\mathrm{A}} = \begin{pmatrix} 1 - 2q_y^2 - 2q_z^2 & 2q_xq_y - 2q_zq_w & 2q_xq_z + 2q_yq_w \\ 2q_xq_y + 2q_zq_w & 1 - 2q_x^2 - 2q_z^2 & 2q_yq_z - 2q_xq_w \\ 2q_xq_z - 2q_yq_w & 2q_yq_z + 2q_xq_w & 1 - 2q_x^2 - 2q_y^2 \end{pmatrix}. \quad \text{(B.22)}$$

A rigid transform $\mathsf{M}_{\mathrm{B}\to\mathrm{A}}$ that transforms Set B's world reference frame to Set A's world reference frame is

formed from $\mathsf{R}_{\mathrm{B}\to\mathrm{A}}$ and the remaining elements of the parameter vector:

$$\mathsf{M}_{\mathrm{B}\to\mathrm{A}} = \begin{pmatrix} \mathsf{R}_{\mathrm{B}\to\mathrm{A}} & \mathbf{t}_{\mathrm{B}\to\mathrm{A}} \\ 0 & 1 \end{pmatrix}, \tag{B.23}$$

where $\mathbf{t}_{\mathrm{B}\to\mathrm{A}} = (x_t, y_t, z_t)^{\mathsf{T}}$.

The 4×4 matrix $\mathsf{M}_{WA\to C1}$ that represents the transform from Set A's world reference frame to Camera 1's reference frame can be written as

$$\mathsf{M}_{WA\to C1} = \begin{pmatrix} m_{A11} & m_{A12} & m_{A13} & m_{A14} \\ m_{A21} & m_{A22} & m_{A23} & m_{A24} \\ m_{A31} & m_{A32} & m_{A33} & m_{A34} \\ m_{A41} & m_{A42} & m_{A43} & m_{A44} \end{pmatrix}. \tag{B.24}$$

The 4×4 matrix $\mathsf{M}_{C2\to WB}$ that represents the transform from Camera 2's reference to Set B's world reference frame can be written as

$$\mathsf{M}_{C2\to WB} = \begin{pmatrix} m_{B11} & m_{B12} & m_{B13} & m_{B14} \\ m_{B21} & m_{B22} & m_{B23} & m_{B24} \\ m_{B31} & m_{B32} & m_{B33} & m_{B34} \\ m_{B41} & m_{B42} & m_{B43} & m_{B44} \end{pmatrix}. \tag{B.25}$$

The transform from Camera 1's reference frame to Camera 2's reference frame is then is computed as follows:

$$\mathsf{M}_{C2\to C1} = (\mathsf{M}_{WA\to C1})(\mathsf{M}_{WB\to WA})(\mathsf{M}_{C2\to WB}). \tag{B.26}$$

This representation of pose is required for computing reprojection errors in the image planes of silhouettes in Set A (since **v** represents a transformation from Set B to Set A).

It is useful to define

$$\mathsf{M}_{C2\to C1} = \begin{pmatrix} r_{11A} & r_{12A} & r_{13A} & t_{xA} \\ r_{21A} & r_{22A} & r_{23A} & t_{yA} \\ r_{31A} & r_{32A} & r_{33A} & t_{zA} \\ 0 & 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \\ a_{41} & a_{42} & a_{43} & a_{44} \end{pmatrix}, \tag{B.27}$$

so that partial derivatives of more than one element of the pose matrices may be specified with a single equation.

It is also useful to define

$$w = {q_1}^2 + {q_2}^2 + {q_3}^2 + {q_4}^2, \tag{B.28}$$

since the expression ${q_1}^2 + {q_2}^2 + {q_3}^2 + {q_4}^2$ occurs frequently.

The remaining equations required for populating the elements of J follow. (These were computed with the aid of the Matlab Symbolic Toolbox.)

$$\frac{\partial a_{ij}}{\partial q_1} =$$
$$\left(m_{Ai1}\left(4\frac{{q_2}^2 q_1}{w^2}+4\frac{{q_3}^2 q_1}{w^2}\right)+m_{Ai2}\left(2\frac{q_2}{w}-4\frac{{q_1}^2 q_2}{w^2}-4\frac{q_3 q_4 q_1}{w^2}\right)+m_{Ai3}\left(2\frac{q_3}{w}-4\frac{{q_1}^2 q_3}{w^2}+4\frac{q_2 q_4 q_1}{w^2}\right)\right)m_{B1j}$$
$$+\left(m_{Ai1}\left(2\frac{q_2}{w}-4\frac{{q_1}^2 q_2}{w^2}+4\frac{q_3 q_4 q_1}{w^2}\right)+m_{Ai2}\left(-4\frac{q_1}{w}+4\frac{{q_1}^3}{w^2}+4\frac{{q_3}^2 q_1}{w^2}\right)+m_{Ai3}\left(-4\frac{q_2 q_3 q_1}{w^2}+2\frac{q_4}{w}-4\frac{{q_1}^2 q_4}{w^2}\right)\right)m_{B2j}$$
$$+\left(m_{Ai1}\left(2\frac{q_3}{w}-4\frac{{q_1}^2 q_3}{w^2}-4\frac{q_2 q_4 q_1}{w^2}\right)+m_{Ai2}\left(-4\frac{q_2 q_3 q_1}{w^2}-2\frac{q_4}{w}+4\frac{{q_1}^2 q_4}{w^2}\right)+m_{Ai3}\left(-4\frac{q_1}{w}+4\frac{{q_1}^3}{w^2}+4\frac{{q_2}^2 q_1}{w^2}\right)\right)m_{B3j} \tag{B.29}$$

$$\frac{\partial a_{ij}}{\partial q_2} =$$
$$\left(m_{Ai1}\left(-4\frac{q_2}{w}+4\frac{{q_2}^3}{w^2}+4\frac{{q_3}^2 q_2}{w^2}\right)+m_{Ai2}\left(-4\frac{q_1 {q_2}^2}{w^2}+2\frac{q_1}{w}-4\frac{q_3 q_4 q_2}{w^2}\right)+m_{Ai3}\left(-4\frac{q_1 q_3 q_2}{w^2}-2\frac{q_4}{w}+4\frac{{q_2}^2 q_4}{w^2}\right)\right)m_{B1j}$$
$$+\left(m_{Ai1}\left(-4\frac{q_1 {q_2}^2}{w^2}+2\frac{q_1}{w}+4\frac{q_3 q_4 q_2}{w^2}\right)+m_{Ai2}\left(4\frac{{q_1}^2 q_2}{w^2}+4\frac{{q_3}^2 q_2}{w^2}\right)+m_{Ai3}\left(2\frac{q_3}{w}-4\frac{{q_2}^2 q_3}{w^2}-4\frac{q_1 q_4 q_2}{w^2}\right)\right)m_{B2j}$$
$$+\left(m_{Ai1}\left(-4\frac{q_1 q_3 q_2}{w^2}+2\frac{q_4}{w}-4\frac{{q_2}^2 q_4}{w^2}\right)+m_{Ai2}\left(2\frac{q_3}{w}-4\frac{{q_2}^2 q_3}{w^2}+4\frac{q_1 q_4 q_2}{w^2}\right)+m_{Ai3}\left(4\frac{{q_1}^2 q_2}{w^2}-4\frac{q_2}{w}+4\frac{{q_2}^3}{w^2}\right)\right)m_{B3j} \tag{B.30}$$

$$\frac{\partial a_{ij}}{\partial q_3} =$$
$$\left(m_{Ai1}\left(4\frac{{q_2}^2 q_3}{w^2}-4\frac{q_3}{w}+4\frac{{q_3}^3}{w^2}\right)+m_{Ai2}\left(-4\frac{q_1 q_2 q_3}{w^2}+2\frac{q_4}{w}-4\frac{{q_3}^2 q_4}{w^2}\right)+m_{Ai3}\left(-4\frac{q_1 {q_3}^2}{w^2}+2\frac{q_1}{w}+4\frac{q_2 q_4 q_3}{w^2}\right)\right)m_{B1j}$$
$$+\left(m_{Ai1}\left(-4\frac{q_1 q_2 q_3}{w^2}-2\frac{q_4}{w}+4\frac{{q_3}^2 q_4}{w^2}\right)+m_{Ai2}\left(4\frac{{q_1}^2 q_3}{w^2}-4\frac{q_3}{w}+4\frac{{q_3}^3}{w^2}\right)+m_{Ai3}\left(-4\frac{q_2 {q_3}^2}{w^2}+2\frac{q_2}{w}-4\frac{q_1 q_4 q_3}{w^2}\right)\right)m_{B2j}$$
$$+\left(m_{Ai1}\left(-4\frac{q_1 {q_3}^2}{w^2}+2\frac{q_1}{w}-4\frac{q_2 q_4 q_3}{w^2}\right)+m_{Ai2}\left(-4\frac{q_2 {q_3}^2}{w^2}+2\frac{q_2}{w}+4\frac{q_1 q_4 q_3}{w^2}\right)+m_{Ai3}\left(4\frac{{q_1}^2 q_3}{w^2}+4\frac{{q_2}^2 q_3}{w^2}\right)\right)m_{B3j} \tag{B.31}$$

$$\frac{\partial a_{ij}}{\partial q_4} =$$
$$\left(m_{Ai1}\left(4\frac{{q_2}^2 q_4}{w^2}+4\frac{{q_3}^2 q_4}{w^2}\right)+m_{Ai2}\left(-4\frac{q_1 q_2 q_4}{w^2}-4\frac{q_3 {q_4}^2}{w^2}+2\frac{q_3}{w}\right)+m_{Ai3}\left(-4\frac{q_1 q_3 q_4}{w^2}+4\frac{q_2 {q_4}^2}{w^2}-2\frac{q_2}{w}\right)\right)m_{B1j}$$
$$+\left(m_{Ai1}\left(-4\frac{q_1 q_2 q_4}{w^2}+4\frac{q_3 {q_4}^2}{w^2}-2\frac{q_3}{w}\right)+m_{Ai2}\left(4\frac{{q_1}^2 q_4}{w^2}+4\frac{{q_3}^2 q_4}{w^2}\right)+m_{Ai3}\left(-4\frac{q_2 q_3 q_4}{w^2}-4\frac{q_1 {q_4}^2}{w^2}+2\frac{q_1}{w}\right)\right)m_{B2j}$$
$$+\left(m_{Ai1}\left(-4\frac{q_1 q_3 q_4}{w^2}-4\frac{q_2 {q_4}^2}{w^2}+2\frac{q_2}{w}\right)+m_{Ai2}\left(-4\frac{q_2 q_3 q_4}{w^2}+4\frac{q_1 {q_4}^2}{w^2}-2\frac{q_1}{w}\right)+m_{Ai3}\left(4\frac{{q_1}^2 q_4}{w^2}+4\frac{{q_2}^2 q_4}{w^2}\right)\right)m_{B3j} \tag{B.32}$$

$$\frac{\partial r_{ijA}}{\partial x_t} = \frac{\partial r_{ijA}}{\partial y_t} = \frac{\partial r_{ijA}}{\partial z_t} = 0 \tag{B.33}$$

$$\begin{pmatrix} \partial t_{xA}/\partial x_t & \partial t_{xA}/\partial y_t & \partial t_{xA}/\partial z_t \\ \partial t_{yA}/\partial x_t & \partial t_{yA}/\partial y_t & \partial t_{yA}/\partial z_t \end{pmatrix} = \begin{pmatrix} m_{A11} & m_{A12} & m_{A13} \\ m_{A21} & m_{A22} & m_{A23} \end{pmatrix}. \tag{B.34}$$

The individual elements of J (see Equation B.3) are computed using the above equations together with the chain rule as shown in the example of Equation B.4.

Appendix C

Polyhedral Models of Stone Data Sets

This appendix shows pictures of the polyhedral models of stones used in this thesis. The polyhedral models are ordered so that similar shapes are nearby. This is achieved using the following approach. Each link between a stone and its four-neighbours is assigned a cost based on the EMD between caliper diameter distributions of the polyhedron. A low-cost combination is computed by randomly selecting pairs of stones, and swapping them if the swap results in a cost reduction. A simulated annealing [51] approach is used so that there is also a slight probability that swaps will be carried out if the swap increases the cost. The probability is controlled by a cooling schedule so that the probability is reduced as the algorithm progresses. This approach allows the algorithm to escape local minima.

The polyhedra have their principal axes aligned with the *x*-, *y*- and *z*-axes, and are viewed from the (1, 1, 1) viewing direction. This helps in visualising the shape of flat or elongated stones in which a large part of the particle may or may not be visible when viewed from an unspecified direction.

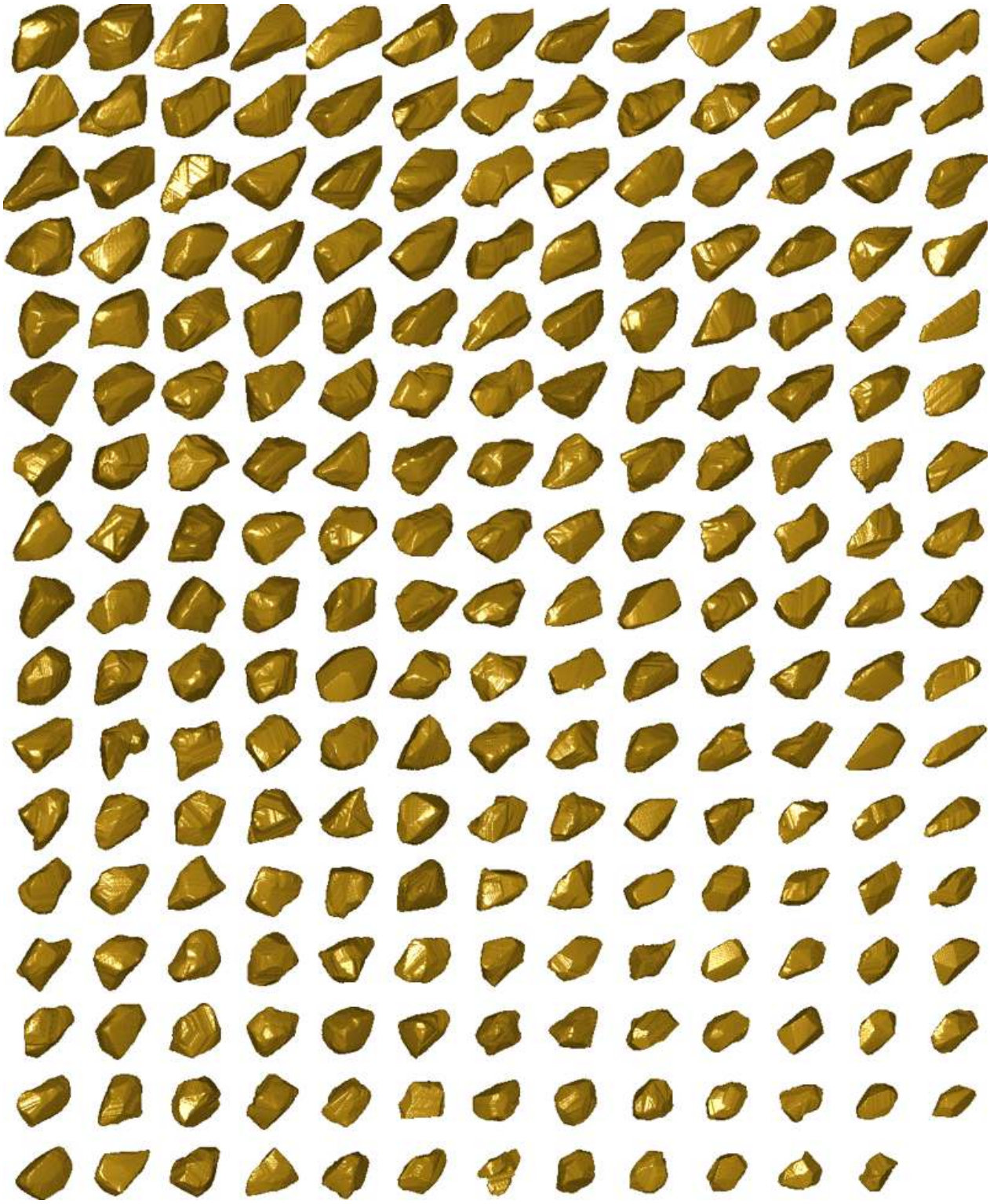


Figure C.1: Refined visual hulls of the data set consisting of 220 pieces of gravel. Each visual hull is formed from 15 views of the stone obtained by merging three runs of 5-view silhouette sets. The 5-view silhouette sets were captured using the mirror setup described in Chapter 4.

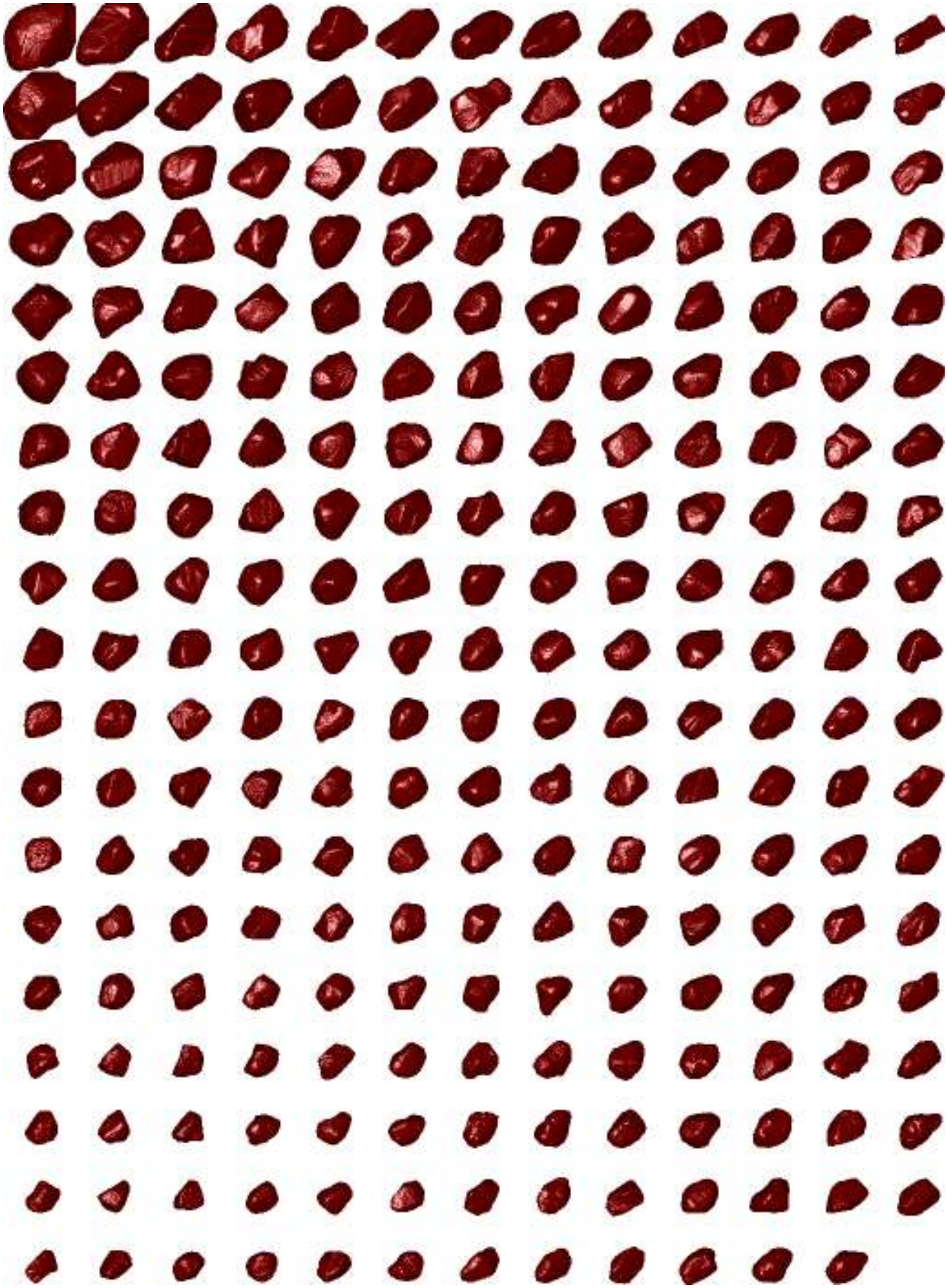


Figure C.2: Refined visual hulls of the data set consisting of 246 garnets. Each visual hull is formed from 30 views of the stone obtained by merging five runs of 6-view silhouette sets. The 6-view silhouette sets were captured using the six-camera setup described in Chapter 5.

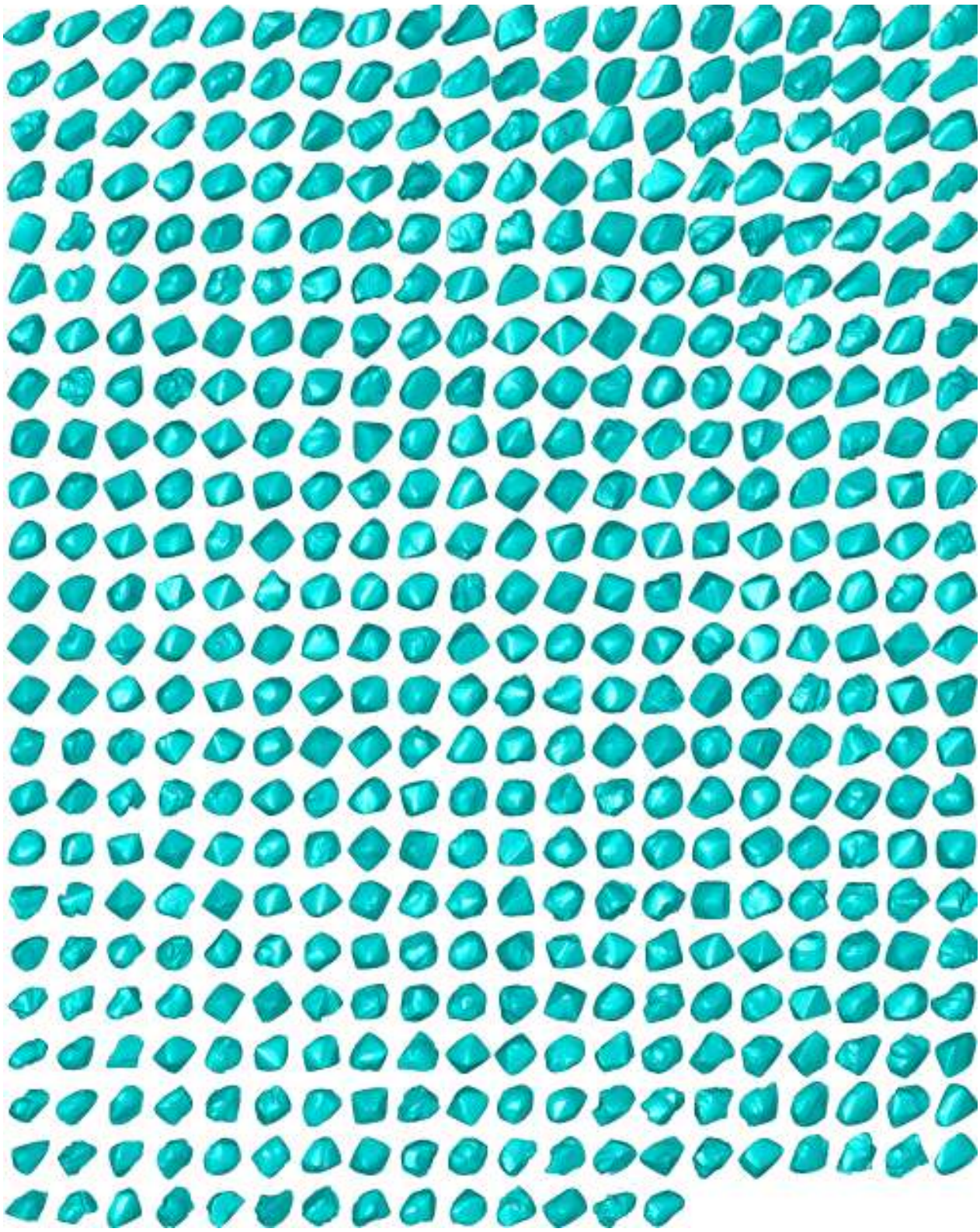


Figure C.3: Refined visual hulls of a portion of the data set consisting of 1426 naturally-occurring (uncut) gemstones. Each visual hull is formed from 60 views of the stone obtained by merging 10 runs of 6-view silhouette sets. The 6-view silhouette sets were captured using the six-camera setup described in Chapter 5. (The other stones from the data set are illustrated on pages 223 and 224.)

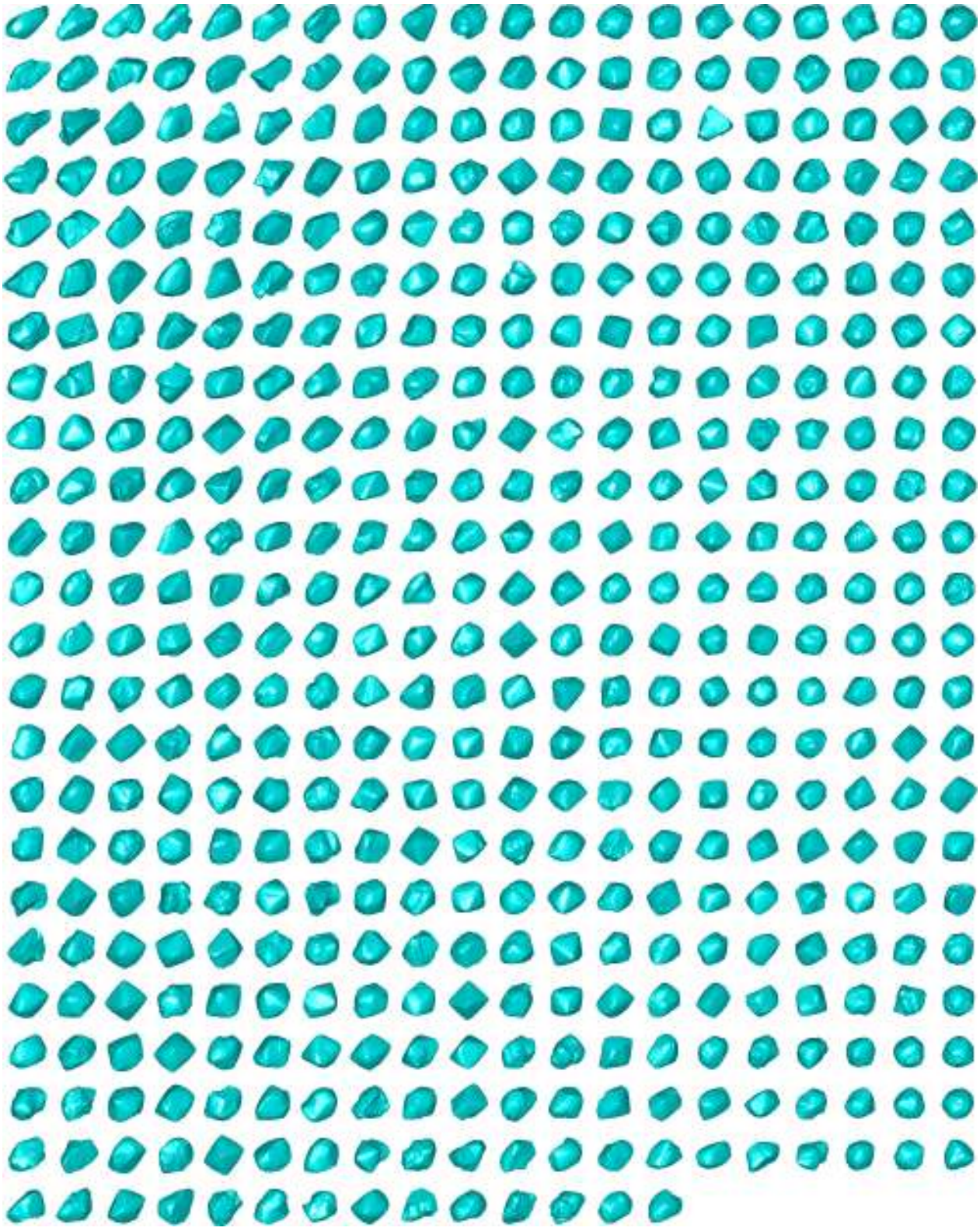


Figure C.4: Refined visual hulls of a portion of the data set consisting of 1426 naturally-occurring (uncut) gemstones. Each visual hull is formed from 60 views of the stone obtained by merging 10 runs of 6-view silhouette sets. The 6-view silhouette sets were captured using the six-camera setup described in Chapter 5. (The other stones from the data set are illustrated on pages 222 and 224.)

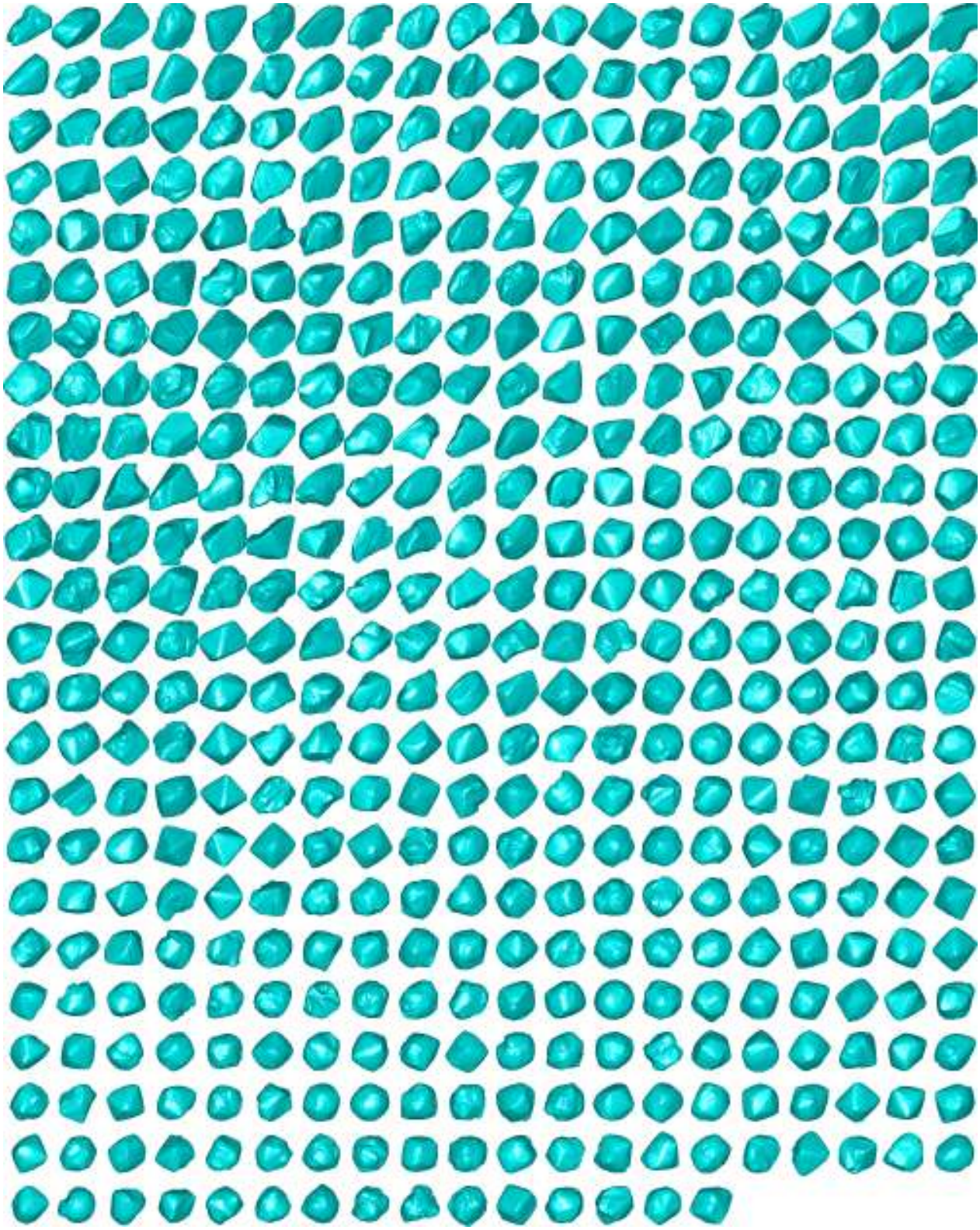


Figure C.5: Refined visual hulls of a portion of the data set consisting of 1426 naturally-occurring (uncut) gemstones. Each visual hull is formed from 60 views of the stone obtained by merging 10 runs of 6-view silhouette sets. The 6-view silhouette sets were captured using the six-camera setup described in Chapter 5. (The other stones from the data set are illustrated on pages 222 and 223.)

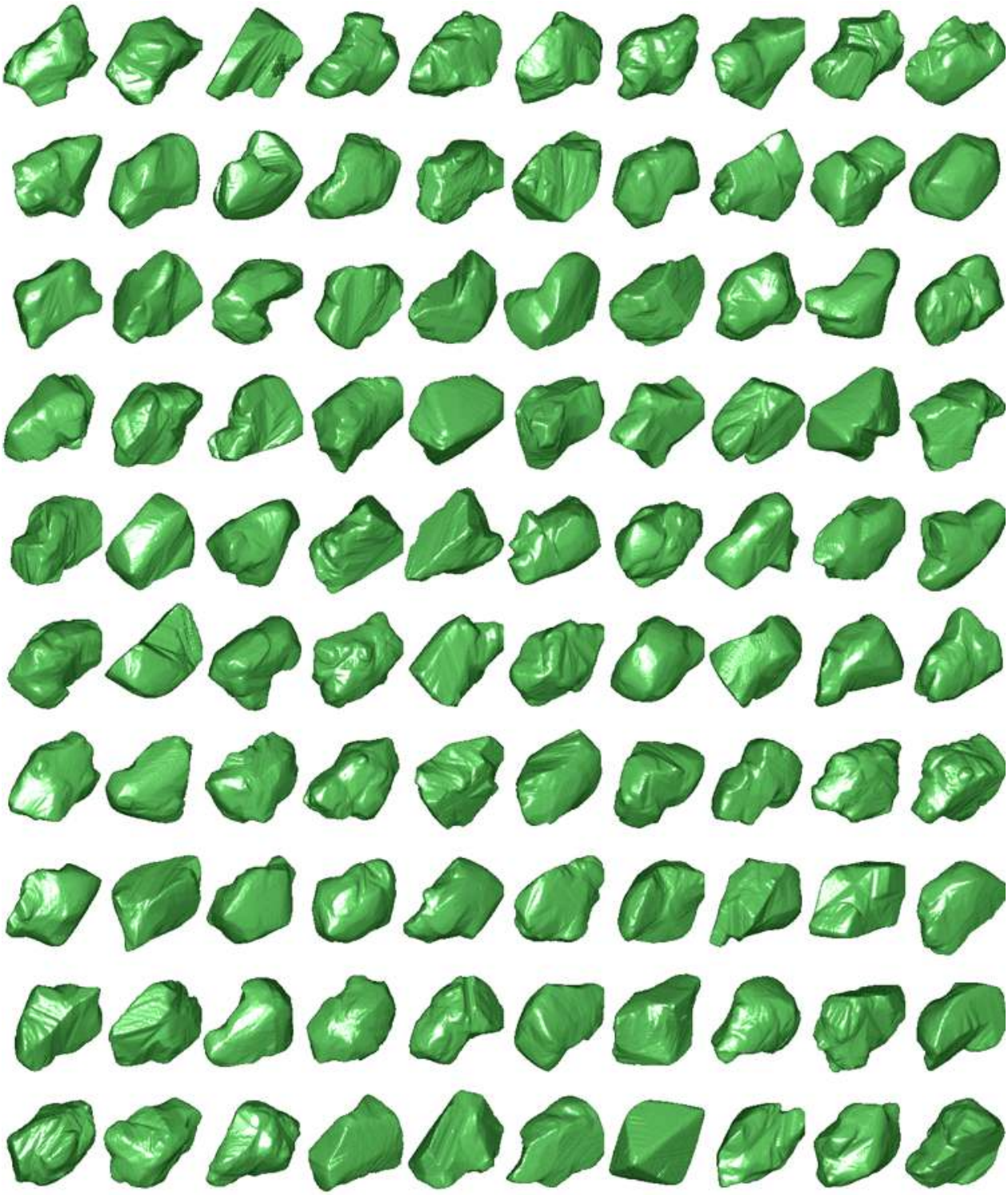


Figure C.6: Synthetic data set of 100 nonconvex stones. The data set is used to investigate matching in Chapter 7.

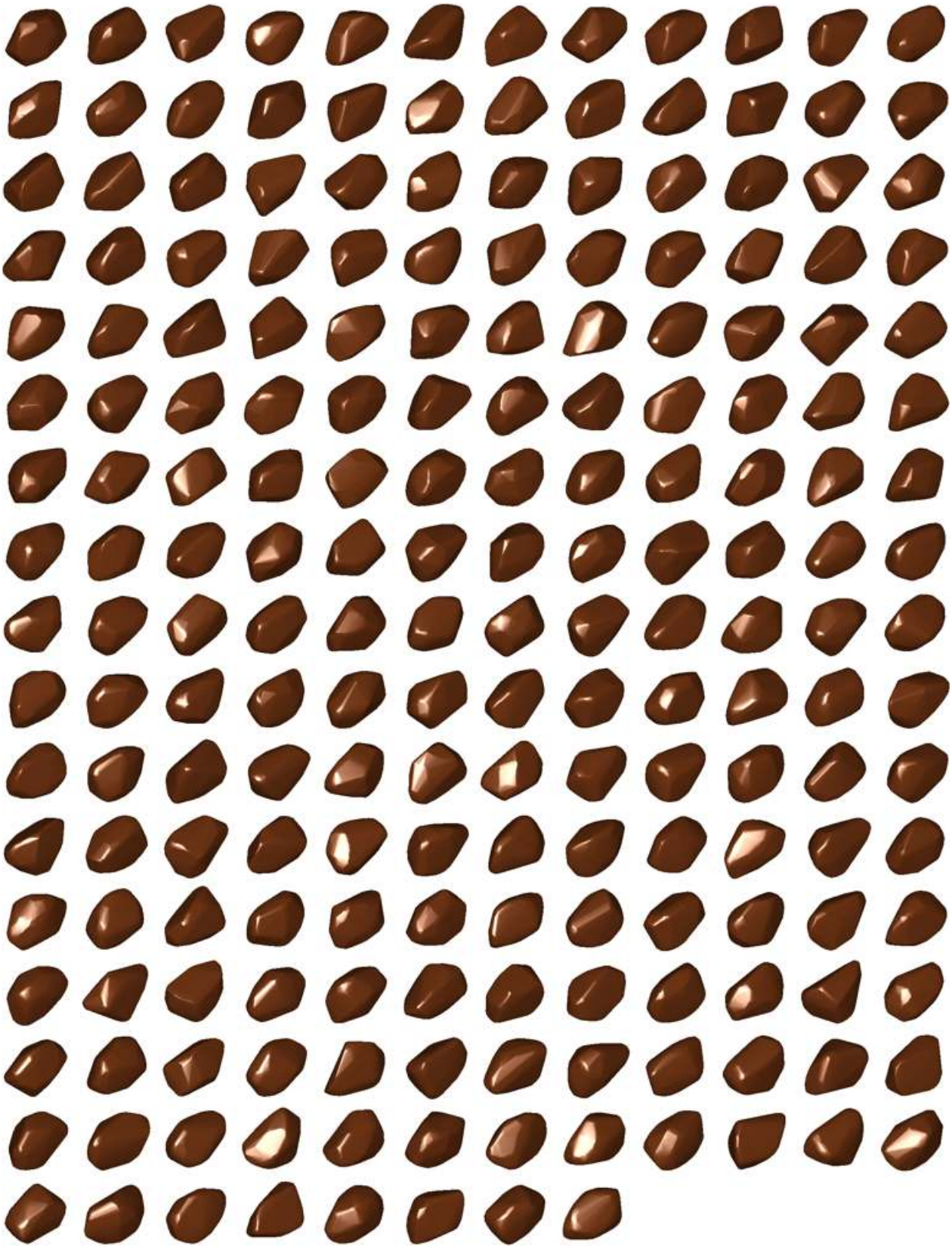


Figure C.7: Synthetic data set of 200 convex stones. The data set is used to investigate matching in Chapter 7.

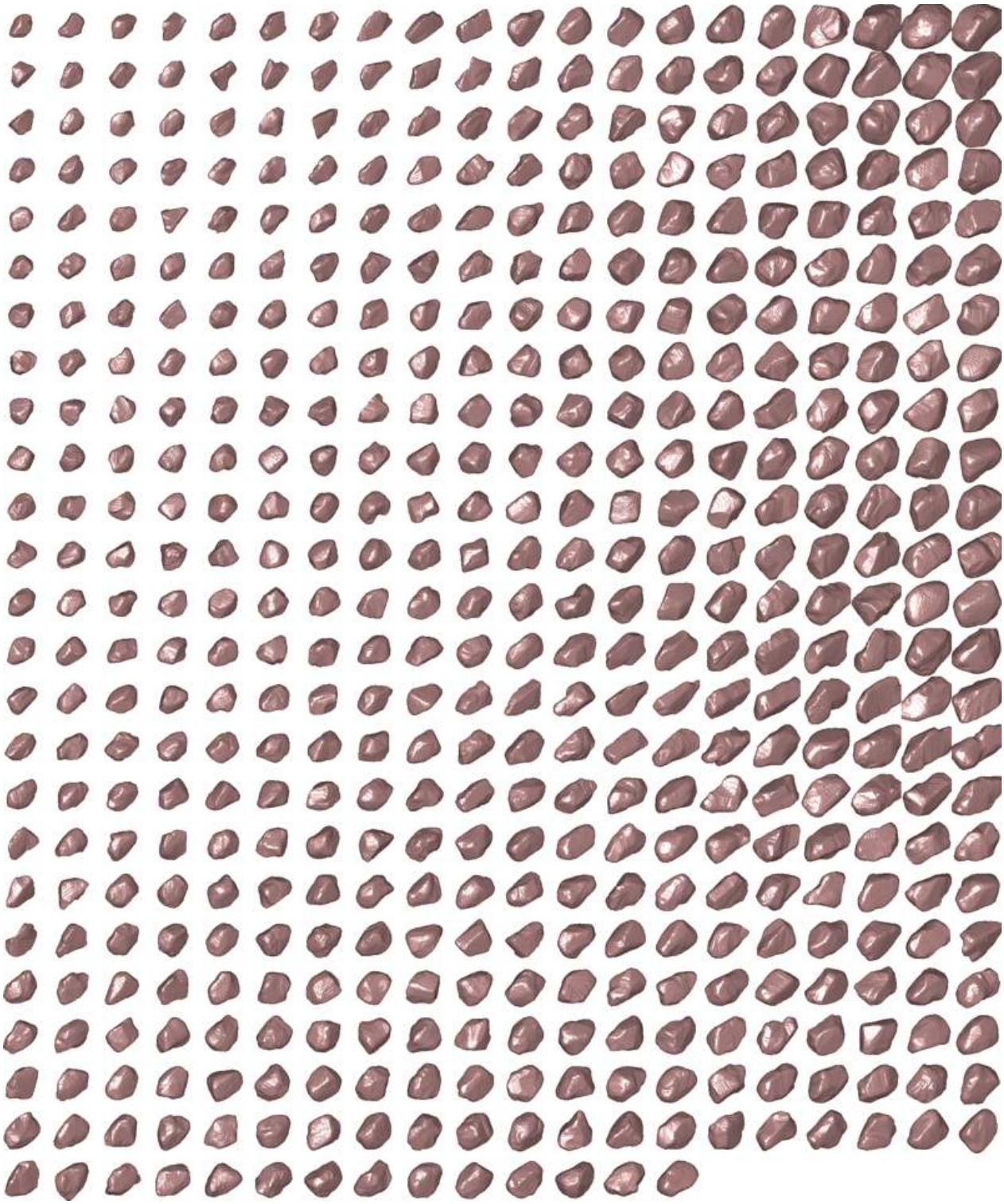


Figure C.8: Refined visual hulls of the data set consisting of 494 garnets used in a sieving experiment described in Chapter 10. Each visual hull is formed from 90 views of the stone obtained by merging 15 runs of 6-view silhouette sets. The 6-view silhouette sets were captured using the six-camera setup described in Chapter 5.

Bibliography

[1] Motilal Agrawal and Larry S. Davis. Camera calibration using spheres: A semi-definite programming approach. In *Proceedings of the 9th International Conference in Computer Vision*, Nice, France, October 2003.

[2] O. Arandjelović. Face recognition from face motion manifolds. First year PhD report, University of Cambridge, September 2004.

[3] K. Åström, R. Cipolla, and P. J. Giblin. Generalised epipolar constraints. *International Journal of Computer Vision*, 1999.

[4] L. Banta, K. Cheng, and J. Zaniewski. Estimation of limestone particle mass from 2D images. *Powder Technology*, 2(24):184–189, June 2003.

[5] C. Bradford Barber, David P. Dobkin, and Hannu Huhdanpaa. The quickhull algorithm for convex hulls. *ACM Transactions on Mathematical Software*, 22(4):469–483, 1996.

[6] P. J. Barrett. The shape of rock particles, a critical review. *Sedimentology*, 27:291–303, 1980.

[7] Bruce G. Baumgart. *Geometric Modeling for Computer Vision*. PhD thesis, Stanford University, 1974. As cited by Lazebnik et al. [76] and others (primary reference inaccessible).

[8] Jules Bloomenthal. An implicit surface polygonizer. In *Graphics Gems IV*, pages 324–349. Academic Press Professional, Inc., San Diego, CA, USA, 1994.

[9] A.L. Bogush, A.V. Tuzikov, and S.A. Sheynin. 3D object reconstruction from non-parallel cross-sections. In *17th International Conference on Pattern Recognition (ICPR)*, 2004.

[10] R M. Bolle, N. K. Ratha, and S. Pankanti. Error analysis of pattern recognition systems – the subsets bootstrap. *Computer Vision and Image Understanding*, 93(1):1–33, January 2004.

[11] A. Bottino and A. Laurentini. Introducing a new problem: Shape-from-silhouette when the relative positions of the viewpoints is unknown. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(11), November 2003.

[12] Andrea Bottino and Aldo Laurentini. What's next? An interactive next best view approach. *Pattern Recognition*, 39(1):126–132, 2006.

[13] Jean-Yves Bouguet. *Visual methods for three-dimensional modeling*. PhD thesis, California Institute of Technology, 1999.

[14] Edmond Boyer. Camera calibration using silhouettes. Technical Report 5559, INRIA, 2005.

[15] Edmond Boyer. On using silhouettes for camera calibration. In *Proceedings of the 7th Asian Conference on Computer Vision (ACCV)*, Hyderabad, India, 2006.

[16] Andrew P. Bradley. The use of the area under the ROC curve in the evaluation of machine learning algorithms. *Pattern Recognition*, 30(7):1145–1159, 1997.

[17] Matthew Brand, Kongbin Kang, and David B. Cooper. Algebraic solution for the visual hull. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2004.

[18] Lars Bretzner. *Multi-Scale Feature Tracking and Motion Estimation*. PhD thesis, Computational Vision and Active Perception Laboratory, KTH, Stockholm, Sweden, 1999.

[19] Kevin Q. Brown. Voronoi diagrams from convex hulls. *Information Processing Letters*, 9:223 – 228, 1979.

[20] Jeffrey W. Bullard, Clarissa Ferraris, Edward J. Garboczi, Nicos Martys, and Paul Stutzman. Virtual cement. In A.K. Bowman and M. Brady, editors, *Innovations in Portland Cement Manufacturing*, chapter 10, pages 1311–1331. Portland Cement Association, 2004.

[21] Benjamin Bustos, Daniel Keim, Dietmar Saupe, Tobias Schreck, and Dejan Vranić. An experimental effectiveness comparison of methods for 3D similarity search. *International Journal on Digital Libraries, Special issue on Multimedia Contents and Management in Digital Libraries*, 6(1):39–54, 2006.

[22] S. Canzar and J. Remy. Shape distributions and protein similarity. In *Proceedings of the German Conference on Bioinformatics*, 2006.

[23] Giorgio Carpaneto and Paolo Toth. Algorithm 548: Solution of the assignment problem. *ACM Transactions on Mathematical Software*, 6(1):104–111, March 1980.

[24] Timothy M. Chan. Approximating the diameter, width, smallest enclosing cylinder, and minimum-width annulus. *International Journal of Computational Geometry and Applications*, 12:67–85, 2002.

[25] Jian-Shiuh Chen, M.K. Chang, and K.Y. Lin. Influence of coarse aggregate shape on the strength of asphalt concrete mixtures. *Journal of the Eastern Asia Society for Transportation Studies*, 6(4):1062 – 1075, 2005.

[26] German K. M. Cheung. *Visual Hull Construction, Alignment and Refinement for Human Kinematic Modeling, Motion Tracking and Rendering*. PhD thesis, Carnegie Mellon University, October 2003.

[27] German K. M. Cheung, Simon Baker, and Takeo Kanade. Visual hull alignment and refinement across time: A 3D reconstruction algorithm combining shape-from-silhouette with stereo. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2003.

[28] Thomas Coleman, Mary Ann Branch, and Andrew Grace. *Optimization Toolbox*. The Mathworks, January 1999.

[29] Simon Collings. *Frontier Points: Theorems and Methods for Computer Vision*. PhD thesis, University of Western Australia, 2001.

[30] Simon Collings, Ryszard Kozera, and Lyle Noakes. Shape recovery of a strictly convex solid from *n* views. In *Proceedings of the 2nd International Conference on Computer Vision and Graphics (ICCVG)*, Warsaw, Poland, September 2004.

[31] Michael Diepenbroek, Alexander Bertholomä, and Hillert Ibbeken. How round is round? A new approach to the topic 'roundness' by Fourier grain shape analysis. *Sedimentology*, 39:411–422, 1992.

[32] G. Drevin and L. Vincent. Granulometric determination of sedimentary rock particle roundness. In *Proc. International Symposium on Mathematical Morphology (ISMM)*, pages 315–325, April 2002.

[33] G. R. Drevin and A. Reyneke. Determining rock particle roundness using the Lorentz curve. In *Proceedings of the Eleventh Annual Symposium of the Pattern Recognition Association of South Africa (PRASA)*, pages 99–104, 2000.

[34] Richard O. Duda, Peter E. Hart, and David G. Stork. *Pattern Classification*. John Wiley and Sons, Inc., 2001.

[35] Bradley Efron. Bayesians, frequentists, and scientists. *Journal of the American Statistical Association*, 100(469), 2004.

[36] Bradley Efron and Robert J. Tibshirani. *An Introduction to the Bootstrap*. Chapman & Hall, 1993.

[37] Robert Ehrlich and Bernhard Weinberg. An exact method for characterization of grain shape. *Journal of Sedimentary Petrology*, 40(1):205–212, March 1970.

[38] S. T. Erdogan, P. N. Quiroga, D. W. Fowler, H. A. Saleh, R. A. Livingston, E. J. Garboczi, P. M. Ketcham, J. G. Hagedorn, and S. G. Satterfield. Three-dimensional shape analysis of coarse aggregates: New techniques for and preliminary results on several different coarse aggregates and reference rocks. *Cement and Concrete Research*, 36:16191627, 2006.

[39] C. Hernández Esteban. *Stereo and Silhouette Fusion for 3D Object Modeling from Uncalibrated Images Under Circular Motion*. PhD thesis, Ecole Nationale Supérieure des Télécommunications, 2004.

[40] O. Faugeras, Q.-T. Luong, and T. Papadopoulo. *The Geometry of Multiple Images: The Laws That Govern the Formation of Multiple Images of a Scene and Some of Their Applications*. MIT Press, 2001.

[41] J. M. R. Fernlund. Image analysis method for determining 3-D shape of coarse aggregate. *Cement and Concrete Research*, 35:1629–1637, 2005.

[42] Joanne M. R. Fernlund. The effect of particle form on sieve analysis: a test by image analysis. *Engineering Geology*, 50:111–124, 1998.

[43] Joanne M. R. Fernlund, Robert W. Zimmerman, and Danica Kragic. Influence of volume/mass on grain-size curves and conversion of image-analysis size to sieve size. *Engineering Geology*, 90:124–137, 2007.

[44] Keith Forbes, Fred Nicolls, Gerhard de Jager, and Anthon Voigt. Shape-from-silhouette with two mirrors and an uncalibrated camera. In *Proceedings of the 9th European Conference on Computer Vision (ECCV)*, Graz, Austria, May 2006.

[45] Keith Forbes, Anthon Voigt, and Ndimi Bodika. An inexpensive, automatic and accurate camera calibration method. In *Proceedings of the Thirteenth Annual Symposium of the Pattern Recognition Association of South Africa (PRASA)*, 2002.

[46] Keith Forbes, Anthon Voigt, and Ndimi Bodika. Using silhouette consistency constraints to build 3D models. In *Proceedings of the Fourteenth Annual Symposium of the Pattern Recognition Association of South Africa (PRASA)*, 2003.

[47] Keith Forbes, Anthon Voigt, and Ndimi Bodika. Visual hulls from single uncalibrated snapshots using two planar mirrors. In *Proceedings of the Fifteenth Annual Symposium of the Pattern Recognition Association of South Africa (PRASA)*, 2004.

[48] Jean-Sébastien Franco and Edmond Boyer. Exact polyhedral visual hulls. In *Proceedings of the British Machine Vision Conference*, pages 329–338, September 2003.

[49] Jean-Sébastien Franco, Marc Lapierre, and Edmond Boyer. Visual shapes of silhouette sets. In *Proceedings of the 3rd International Symposium on 3D Data Processing, Visualization and Transmission*, 2006.

[50] J. D. Frost and J. S. Lai. Digital analysis of aggregate particle shape. In *Proceedings of 4th Annual Symposium, Center for Aggregates Research*, 1996.

[51] Mark Galassi, Jim Davies, James Theiler, Brian Gough, Gerard Jungman, Michael Booth, and Fabrice Rossi. *Gnu Scientific Library Reference Manual*. Network Theory Ltd, January 2002. Also available at `http://sources.redhat.com/gsl`.

[52] Edward J. Garboczi, Jeffrey W. Bullard, and Dale P. Bentz. Virtual testing of cement and concrete—USA 2004. *Concrete International*, 26(12):33–37, December 2004.

[53] J.M. Gluckman and S.K. Nayar. Planar catadioptric stereo: Geometry and calibration. In *Proceedings of the Conference on Computer Vision and Pattern Recognition (CVPR)*, 1999.

[54] Aldo A. Grattarola. Volumetric reconstruction from object silhouettes: A regularization procedure. *Signal Processing*, 27(1):27–35, 1992.

[55] J. Y. Han and K. Perlin. Measuring bidirectional texture reflectance with a kaleidoscope. In *Proceedings of SIGGRAPH*, 2003.

[56] Mei Han and Takeo Kanade. Creating 3D models with uncalibrated cameras. In *Proceeding of IEEE Computer Society Workshop on the Application of Computer Vision (WACV2000)*, December 2000.

[57] Sariel Har-Peled. A practical approach for computing the diameter of a point set. In *SCG '01: Proceedings of the Seventeenth Annual Symposium on Computational Geometry*, pages 177–186, New York, NY, USA, 2001. ACM Press.

[58] R. I. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, second edition, 2004.

[59] Lauren E. Hassler and Ellen A. Cowan. Characteristics of ice-rafted pebbles from the continental rise sediment drifts west of the Antarctic peninsula. In P. F. Barker, A. Camerlenghi, G. D. Acton, and A. T. S. Ramsay, editors, *Proceedings of the Ocean Drilling Program, Scientific Results*, volume 178, 2001.

[60] C. Hernández, F. Schmitt, and R Cipolla. Silhouette coherence for camera calibration under circular motion. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(2):343–349, February 2007.

[61] Berthold Klaus Paul Horn. *Robot Vision*. The MIT Electrical Engineering and Computer Science Series. The MIT Press, 1986.

[62] B. Hu, C. M. Brown, and R. C. Nelson. Multiple-view 3-D reconstruction using a mirror. Technical Report TR863, Computer Science Dept., University of Rochester, May 2005.

[63] Po-Hao Huang and Shang-Hon Lai. Contour-based structure from reflection. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2006.

[64] Cheuk Yiu Ip, Daniel Lapadat, Leonard Sieger, and William C. Regli. Using shape distributions to compare solid models. In *Proceedings of the Symposium on Solid Modeling and Applications*, 2002.

[65] David Jacobs, Peter Belhumeur, and Ian H. Jermyn. Judging whether multiple silhouettes can come from the same object. In *Proceedings of the 4th International Workshop on Visual Form*, pages 532–541, Capri, Italy, 2001.

[66] A. K. Joshi and R. Bajcsy. The problem of naming shapes: Vision-language interface. In *Proceedings of the 2nd Workshop on Theoretical Issues in Natural Language Processing*, July 1978.

[67] Manfred Kaessens. How does your fruit shape up? *Photonics Showcase*, pages 13–14, November 2006.

[68] Kongbin Kang. *Three-Dimensional Free Form Surface Reconstruction from Occluding Contours in A Sequence Images or Video*. PhD thesis, Brown University, 2004.

[69] J. J. Koenderink. What does the occluding contour tell us about solid shape? *Perception*, 13:321–330, 1984.

[70] W. C. Krumbein. Measurement and geological significance of shape and roundness of sedimentary particles. *Journal of Sedimentary Petrology*, 11:64–72, 1941.

[71] F. Lanaro and P. Tolppanen. 3D characterization and analysis of particle shape using x-ray microtomography (XMT). *Powder Technology*, 154:61–69, 2005.

[72] Tony Lau. Using 2-D projections to characterize 3-D particles. Master's thesis, Department of Civil and Environmental Engineering, University of California, 2003.

[73] Aldo Laurentini. The visual hull concept for silhouette-based image understanding. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 16(2):150–162, 1994.

[74] Aldo Laurentini. How far 3D shapes can be understood from 2D silhouettes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 17(2):188–195, February 1995.

[75] Svetlana Lazebnik. Projective visual hulls. Master's thesis, University of Illinois at Urbana-Champaign, 2002.

[76] Svetlana Lazebnik, Edmond Boyer, and Jean Ponce. On computing exact visual hulls of solids bounded by smooth surfaces. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2001.

[77] Svetlana Lazebnik, Amit Sethi, Cordelia Schmid, David Kriegman, Jean Ponce, and Martial Hebert. On pencils of tangent planes and the recognition of smooth 3D shapes from silhouettes. In *Proceedings of the European Conference on Computer Vision*, 2002.

[78] Alfred A. Levinson and James E. Shigley. White House conference on "conflict" gemstones. *Gems & Gemology*, 37(1), 2001.

[79] J. Li, C. Webb, S. S. Pandiella, and G. M. Campbell. Discrete particle motion on sieves—a numerical study using the DEM simulation. *Powder Technology*, 133:190–202, 2003.

[80] C. Liang and K.-Y. K. Wong. Complex 3D shape recovery using a dual-space approach. In *Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition*, volume 2, pages 878–884, June 2005.

[81] Sheue-ling Lien and James T. Kajiya. A symbolic method for calculating the integral properties of arbitrary nonconvex polyhedra. *IEEE Computer Graphics and Applications*, 4(10):35–41, October 1984.

[82] C. L. Lin and J. D. Miller. 3D characterization of coarse aggregates. *Engineering Geology*, 65:17–30, 2002.

[83] David A. Lindsey and John F. Shary. Field measures of gravel quality in the South Platte River north of Denver, Colorado: A pilot study, 2003. Open-File Report 97-451, U.S. Department of the Interior.

[84] William E. Lorensen and Harvey E. Cline. Marching cubes: A high resolution 3D surface construction algorithm. *Computer Graphics*, 21(4):163–169, 1987.

[85] Alexander Lyons, Adam Baumberg, and Aaron Kotcheff. SAVANT: A new efficient approach to generating the visual hull. In *Short Paper Proceedings, Eurographics*, Saarbrücken, Germany, 2002.

[86] Norbert H. Maerz. Technical and computational aspects of the measurement of aggregate shape by digital image analysis. *Journal of Computing in Civil Engineering*, 18(1):10–18, January 2004.

[87] Norbert H. Maerz and Mike Lusher. Measurement of flat and elongation of coarse aggregate using digital image processing. In *Proceedings of the Transportation Research Board 80th Annual Meeting*, pages 2–14, 2001.

[88] David Marr. *Vision*. W. H. Freeman and Company, 1980.

[89] Alvin Martin, George Doddington, Terri Kamm, Mark Ordowski, and Mark Przybocki. The DET curve in assessment of detection task performance. In *Proceedings of Eurospeech*, pages 1895–1898, Rhodes, Greece, 1997.

[90] Wojciech Matusik. Image-based visual hulls. Master's thesis, Massachusetts Institute of Technology, February 2000.

[91] Wojciech Matusik, Chris Buehler, and Leonard McMillan. Polyhedral visual hulls for real-time rendering. In *Proceedings of Twelfth Eurographics Workshop on Rendering*, pages 115–125, June 2001.

[92] Avraham A. Melkman. On-line construction of the convex hull of a simple polyline. *Information Processing Letters*, 25(1):11–12, April 1987.

[93] P. R. S. Mendonça, K.-Y. K. Wong, and R. Cipolla. Epipolar geometry from profiles under circular motion. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(6):604–616, June 2001.

[94] Brian Mirtich. Fast and accurate computation of polyhedral mass properties. *Journal of Graphics Tools*, 1(2):31–50, 1996.

[95] J. J. Moré, B. S. Garbow, and K. E. Hillstrom. User's guide for MINPACK–1. Technical report, Argonne National Laboratory, Argonne, IL, USA, 1980. Tech. Rep. ANL–80–74.

[96] T. Moriya, F. Beniyama, K. Matsumoto, T. Minakawa, K. Utsugi, and H. Takeda. Properties of three shadows on a plane. In *Proceedings of the 12th International Conference in Central Europe on Computer Graphics, Visualization and Computer Vision*, 2004.

[97] Adem Yaşar Mülayim, Ulaş Yilmaz, and Volkan Atalay. Silhouette-based 3D model reconstruction from multiple images. *IEEE Transactions on Systems, Man and Cybernetics*, 33(4):582–591, 2003.

[98] Lars Mundermann, Stefano Corazza, Ajit M. Chaudhari, Eugene J. Alexander, and Thomas P. Andriacchi. Most favorable camera configuration for a shape-from-silhouette markerless motion capture system for biomechanical analysis. In *Proceedings of SPIE, vol. 5665, Videometrics VIII*, 2005.

[99] W. Niem and R. Buschmann. Automatic modelling of 3D natural objects from multiple views. In *European Workshop on Combined Real and Synthetic Image Processing for Broadcast and Video Production*, Hamburg, Germany, November 1994.

[100] J. Nyssen, J. Poesen, J. Moeyersons, J. Deckers, and Mitiku Haile. Processes and rates of rock fragment displacement on cliffs and scree slopes in an amba landscape, Ethiopia. *Geomorphology*, 81:265–275, November 2006.

[101] Takayuki Okatani and Koichiro Deguchi. Recovering camera motion from image sequence based on registration of silhouette cones. In *Proceedings of the 6th IAPR Workshop on Machine Vision Applications (MVA2000)*, pages 451–454, 2000.

[102] Takayuki Okatani and Koichiro Deguchi. Recovering camera motion from image sequence based on registration of silhouette cones: Shape from silhouette using a mobile camera with a gyro sensor. In *Proceedings of the 6th IAPR Workshop on Machine Vision Applications*, pages 451–454, 2000.

[103] Robert Osada, Thomas Funkhouser, Bernard Chazelle, and David Dobkin. Matching 3D models with shape distributions. In *Proceedings of the Third International Conference on Shape Modeling and Applications*, Genova, Italy, May 2001.

[104] Robert Osada, Thomas Funkhouser, Bernard Chazelle, and David Dobkin. Shape distributions. *ACM Transactions on Graphics*, 21(4), 2002.

[105] N. Otsu. A threshold selection method from gray-level histograms. *IEEE Transactions on Systems, Man, and Cybernetics*, 9(1):62–66, 1979.

[106] P. Jonathon Phillips, Hyeonjoon Moon, Syed A. Rizvi, and Patrick J. Rauss. The FERET evaluation methodology for face-recognition algorithms. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(10):1090–1104, October 2000.

[107] Mathew Price and Garry Morrison. Validating rigid body simulation of real particle shapes using pose estimation from high-speed video. In *Proceedings of the 4th International Conference on Discrete Element Methods*, Brisbane, Australia, August 2007.

[108] C. Rao and E. Tutumluer. Determination of volume of aggregates: New image analysis approach. *Transportation Research Record*, 1721:73–80, 2000.

[109] Chetana B. Rao. *Development of 3-D Image Analysis Techniques to Determine Shape and Size Properties of Coarse Aggregate*. PhD thesis, University of Illinois at Urbana-Champaign, 2001.

[110] David Renaudie, David J. Kriegman, and Jean Ponce. Duals, invariants, and the recognition of smooth objects from their occluding contours. In *Proceedings of the 7th European Conference on Computer Vision (ECCV)*, pages 784–798, 2000.

[111] Taleb Mustafa Al Rousan. *Characterization of Aggregate Shape Properties Using a Computer Automated System*. PhD thesis, Texas A&M University, 2004.

[112] Y. Rubner, C. Tomasi, and L. Guibas. The earth mover's distance as a metric for image retrieval. Technical report, Department of Computer Science, Stanford University, September 1998. Technical Report STAN-CS-TN-98-86.

[113] Y. Rubner, C. Tomasi, and L. Guibas. The earth mover's distance as a metric for image retrieval. *International Journal of Computer Vision*, 40(2):99121, 2000.

[114] Robert Sablatnig, Srdan Tosovic, and Martin Kampel. Next view planning for shape from silhouette. In Ondrej Drbohlav, editor, *Proceedings of the 8th Computer Vision Winter Workshop*, pages 77–82, Valtice, Czech Republic, 2003. Czech Pattern Recognition Society.

[115] M. Sainz, N. Bagherzadeh, and A. Susin. Recovering 3D metric structure and motion from multiple uncalibrated cameras. In *Proceedings of IEEE Conference on Information Technology: Coding and Computing*, 2002.

[116] Elmar Schömer, Jürgen Sellen, Marek Teichmann, and Chee Yap. Smallest enclosing cylinders. *Algorithmica*, 27:170–186, 2000.

[117] H. Schwarcz and K. Shane. Measurement of particle shape by Fourier analysis. *Sedimentology*, 13:213231, 1969.

[118] Guruprasd Shivaram and Guna Seetharaman. A simple and fast technique for camera calibration using spheres as test objects. In *Proceedings of the IEEE International Workshop on 3-D Image Analysis and Synthesis*, Alpbach, Austria, July 1998.

[119] J. S. Sims, J. G. Hagedorn, P. M. Ketcham, S. G. Satterfield, T. J. Griffin, W. L. George, H. A. Fowler, B. A. Ende, H. K. Hung, R. B. Bohn, J. E. Koontz, N. S. Martys, C. E. Bouldin, J. A. Warren, D. L. Feder, C. W. Clark, B. J. Filla, and J. E. Devaney. Accelerating scientific discovery through computation and visualization. *NIST Journal of Research*, 107(3):223–245, 2002.

[120] Sudipta N. Sinha and Marc Pollefeys. Visual-hull reconstruction from uncalibrated and unsynchronized video streams. In *Proceedings of the Second International Symposium on 3D Data Processing, Visualization and Transmission*, pages 349–356, 2004.

[121] S. Sullivan and J. Ponce. Automatic model construction and pose estimation from photographs using triangular splines. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(10):1091–1096, 1998.

[122] Tomáš Svoboda, Hanspeter Hug, and Luc Van Gool. ViRoom – low cost synchronized multi-camera system and its self-calibration. In *Pattern Recognition, 24th DAGM Symposium*, Zurich, Switzerland, September 2002.

[123] R. Szeliski. Rapid octree construction from image sequences. *Computer Vision, Graphics and Image Processing*, 58(1):2332, July 1993.

[124] Johan W. H. Tangelder and Remco C. Veltkamp. A survey of content based 3D shape retrieval methods. In *Proceedings of the International Conference on Shape Modeling and Applications*, pages 145–156, 2004.

[125] Marco Tarini, Marco Callieri, Claudio Montani, Claudio Rocchini, Karin Olsson, and Therese Persson. Marching intersections: An efficient approach to shape-from-silhouette. In Bernd Girod, Günther Greiner, Heinrich Niemann, and Hans-Peter Seidel, editors, *Proceedings of the Conference on Vision, Modeling, and Visualization*, pages 255–262, November 2002.

[126] Michael Taylor. Using multiple 2-D projections to characterize 3-D irregular particles. In *Proceedings of the 12th Annual Symposium of the International Center for Aggregates Research*, 2004.

[127] Michael A. Taylor. Quantitative measures for shape and size of particles. *Powder Technology*, 124(24):94–100, 2002.

[128] H. Teramoto and G. Xu. Camera calibration by a single image of balls: From conics to the absolute conic. In *Proceedings of the 5th Asian Conference on Computer Vision*, pages 499–506, 2002.

[129] Carlo Tomasi and Takeo Kanade. Shape and motion from image streams: a factorization method. Technical Report CMU-CS-91-132, Computer Science Department, Carnegie Mellon University, Pittsburgh, PA, April 1991.

[130] R. Y. Tsai. A versatile camera calibration technique for high-accuracy 3D machine vision metrology using off-the-shelf TV cameras and lenses. *IEEE Journal of Robotics and Automation*, 3(4):232–344, August 1987.

[131] L. Uthus, I. Hoff, and I. Horvli. Evaluation of grain shape characterization methods for unbound aggregates. In *Proceedings of the Seventh International Conference on the Bearing Capacity of Road, Railways and Airfields (BCRA)*, 2005.

[132] G. Vogiatzis, P. Torr, and R. Cipolla. Multi-view stereo via volumetric graph-cuts. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2005.

[133] Hakon Wadell. Volume, shape and roundness of rock particles. *Journal of Geology*, 40:443–451, 1932.

[134] Daphna Weinshall and Carlo Tomasi. Linear and incremental acquisition of invariant shape models from image sequences. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 17(5):512–517, 1995.

[135] Emo Welzl. Smallest enclosing disks (balls and ellipsoids). *Lecture Notes in Computer Science*, 555:359–370, 1991.

[136] J. Wierchowiec. Morphology and chemistry of placer gold grains – indicators of the origin of placers: an example from the East Sudetic Foreland, Poland. *Acta Geologica Polonica*, 52(4):563576, 2002.

[137] Jochen Wingbermühle. Robust registration of coarse binary volume models. In *Proceedings of the International Conference on Pattern Recognition*, volume 1, Barcelona, Spain, September 2000.

[138] K.-Y. K. Wong. *Structure and Motion from Silhouettes*. PhD thesis, University of Cambridge, October 2001.

[139] K.-Y. K. Wong and R. Cipolla. Reconstruction of sculpture from its profiles with unknown camera positions. *IEEE Transactions on Image Processing*, 13(3):381–389, 2004.

[140] Gang Xu, Yuki Inoue, and Hirohisa Teramoto. Determining camera intrinsic and extrinsic parameters from multiple images of multiple balls. In *Proceedings of the Workshop on Science of Computer Vision*, September 2002.

[141] Hirotake Yamazoe, Akira Utsumi, and Shinji Abe. Multiple camera calibration with bundled optimization using silhouette geometry constraints. In *Proceedings of the 18th International Conference on Pattern Recognition (ICPR)*, 2006.

[142] Anna Yershova and Steven M. LaValle. Deterministic sampling methods for spheres and SO(3). In *Proceedings of the IEEE International Conference on Robotics and Automation*, 2004.

[143] C. Zhang and T. Chen. Efficient feature extraction for 2D/3D objects in mesh representation. In *Proceedings of the 8th International Conference on Image Processing (ICIP)*, October 2001.

[144] T. Zingg. Beitrage zur schotteranalyse. [Contribution to crushed stone analysis.] *Schweizerische Mineralogische und Petrographische Mitteilungen* [Swiss Bulletin of Mineralogy and Petrology], 15:38–140, 1935. As cited by Barettt [6] and others (primary reference inaccessible).