



US00RE36656E

[54] **GENERALIZED SHAPE
AUTOCORRELATION FOR SHAPE
ACQISITION AND RECOGNITION**

[75] Inventors: **Andrea Califano**, New York, N.Y.;
Rakesh Mohan, Stamford, Conn.

[73] Assignee: **International Business Machines
Corporation**, Armonk, N.Y.

[21] Appl. No.: **08/719,496**

[22] Filed: **Sep. 25, 1996**

Related U.S. Patent Documents

Reissue of:

[64] Patent No.: **5,351,310**
Issued: **Sep. 27, 1994**
Appl. No.: **08/162,966**
Filed: **Dec. 8, 1994**

U.S. Applications:

[63] Continuation of application No. 07/705,037, May 21, 1991,
abandoned.

[51] Int. Cl.⁷ **G06K 9/48**; G06K 9/46;
G06K 9/36

[52] U.S. Cl. **382/199**; 382/203; 382/276

[58] Field of Search 382/203, 199,
382/276, 195, 204, 190; 348/125, 134,
135, 141

References Cited

U.S. PATENT DOCUMENTS

3,069,654 12/1962 Hough et al. 382/41
4,110,736 8/1978 Kono 382/203
4,115,761 9/1978 Ueda et al. 382/203
4,115,805 9/1978 Morton 382/305
4,183,013 1/1980 Agrawala et al. 382/173
4,388,610 6/1983 Tsunekawa 382/113
4,704,694 11/1987 Czerniejewski 364/513

4,747,153 5/1988 Kouno et al. 382/203
4,748,674 5/1988 Freeman 382/157
4,783,829 11/1988 Miyakawa et al. 382/199
4,817,183 3/1989 Sparrow 382/125
4,845,765 7/1989 Juvin et al. 382/153
4,933,865 6/1990 Yamamoto et al. 382/203
4,969,201 11/1990 Takasaki et al. 382/199

OTHER PUBLICATIONS

Califano et al., "Generalized Shape Autocorrelation",
AAAI-90 Proceeding Eighth National Conference on Arti-
ficial Intelligence, vol. 2, Jul. 29-Aug. 3, 1990 pp.
1067-1073.

Bolle et al., "A Complete and Scalable Architecture for 3D
Model-Based Vision", Proceedings, 5th IEEE International
Symposium on Intelligent Control 1990, vol. 1, Sep. 5-7,
1990, pp. 212-219.

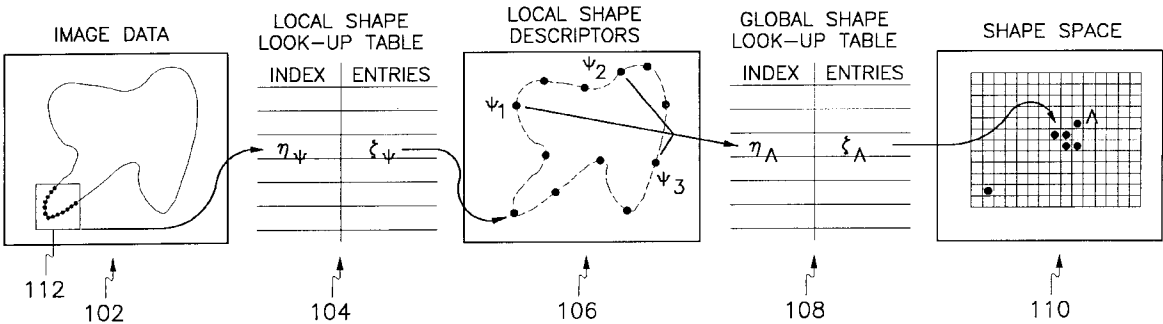
Y. Lamdan and H. J. Wolfson, "Geometric Hashing: A
General and Efficient Model-Based Recognition Scheme,"
Proc. of IEEE Second Int'l Conf. on Computer Vision, pp.
238-249, Tampa, Florida, Dec., 1988.

Primary Examiner—Bipin Shalwala
Attorney, Agent, or Firm—Jay P. Sbröllini

[57] **ABSTRACT**

The invention provides automatic acquisition and recognition of complex visual shapes within images. During an acquisition phase, models are derived from interest points acquired from a target shape. The models are stored in and can be retrieved from a lookup table via high dimension indices. When an image is inputted, triplets of interest points in the image are used to compute local shape descriptors, which describe the geometry of local shapes in the image. In turn, triplets of local shape descriptors are used to compute high dimension indices. These indices are used for accessing the lookup table having the models. The models are used for the automatic recognition of target shapes.

10 Claims, 9 Drawing Sheets



100

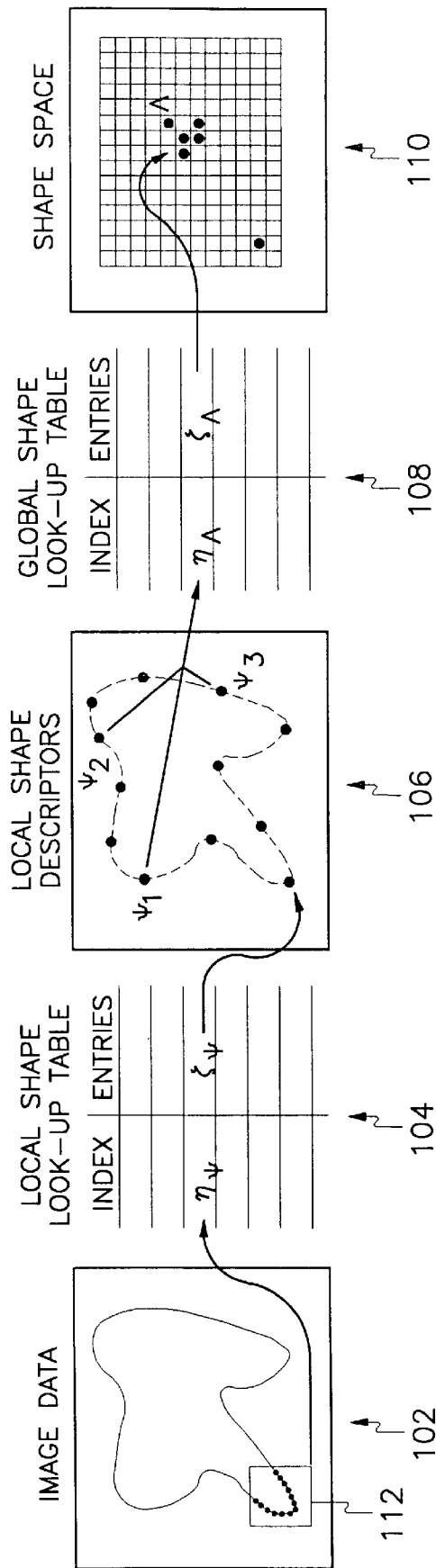


Fig. 1

200

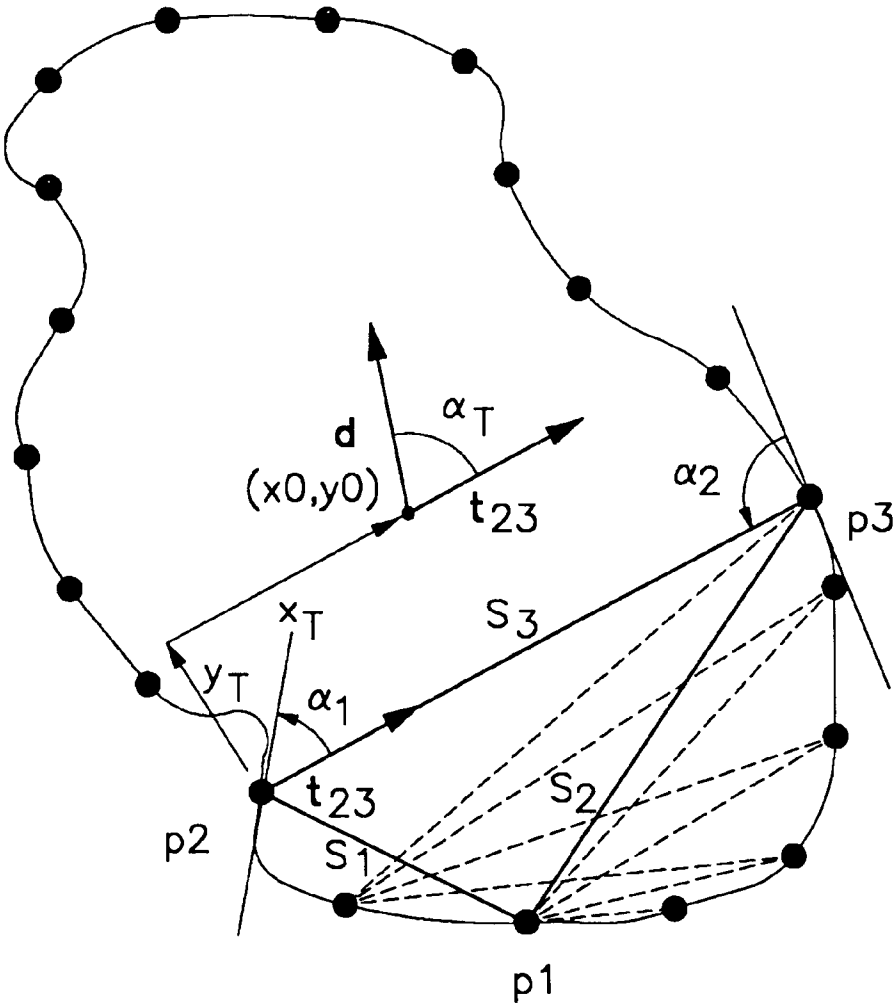


Fig. 2

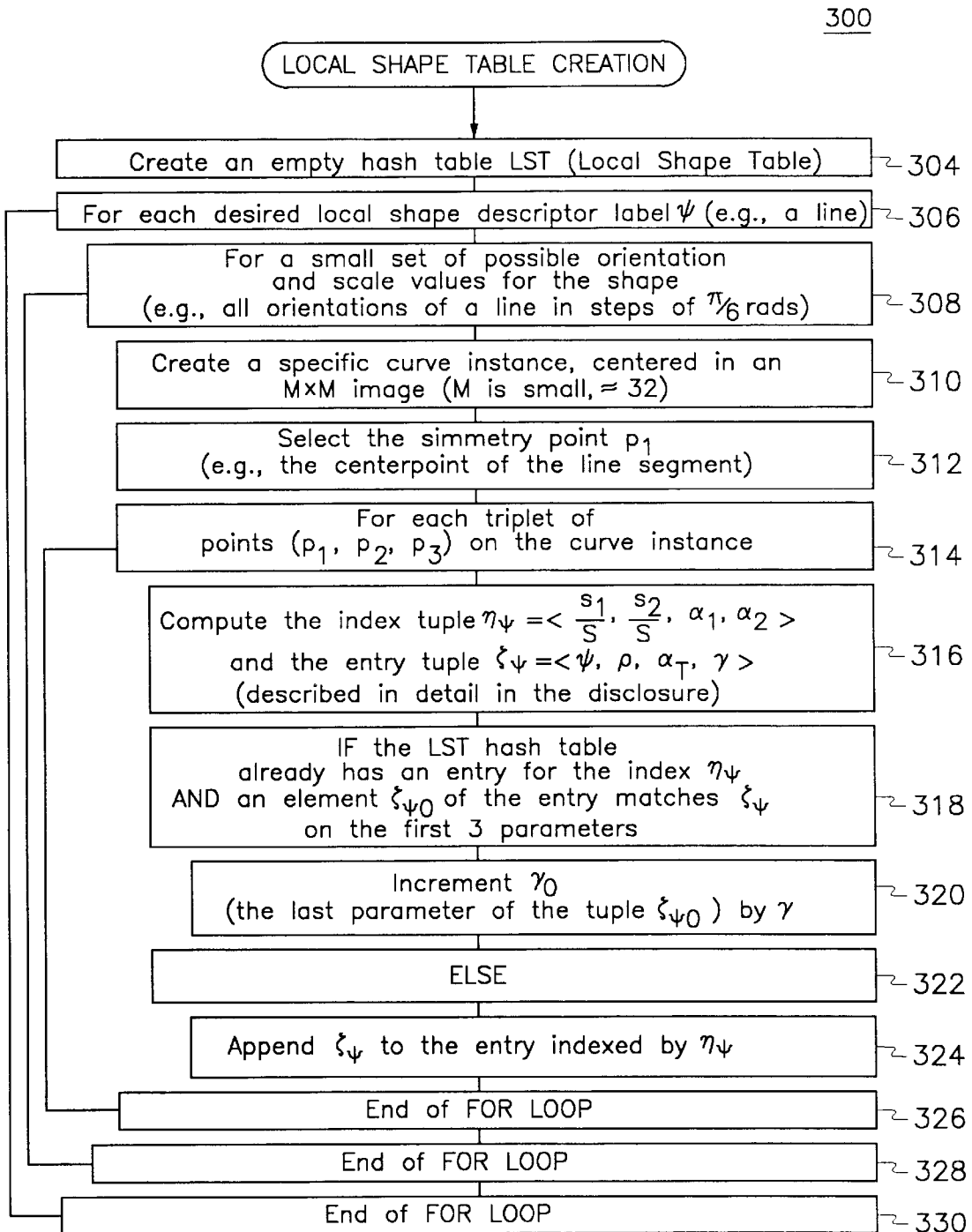


Fig. 3

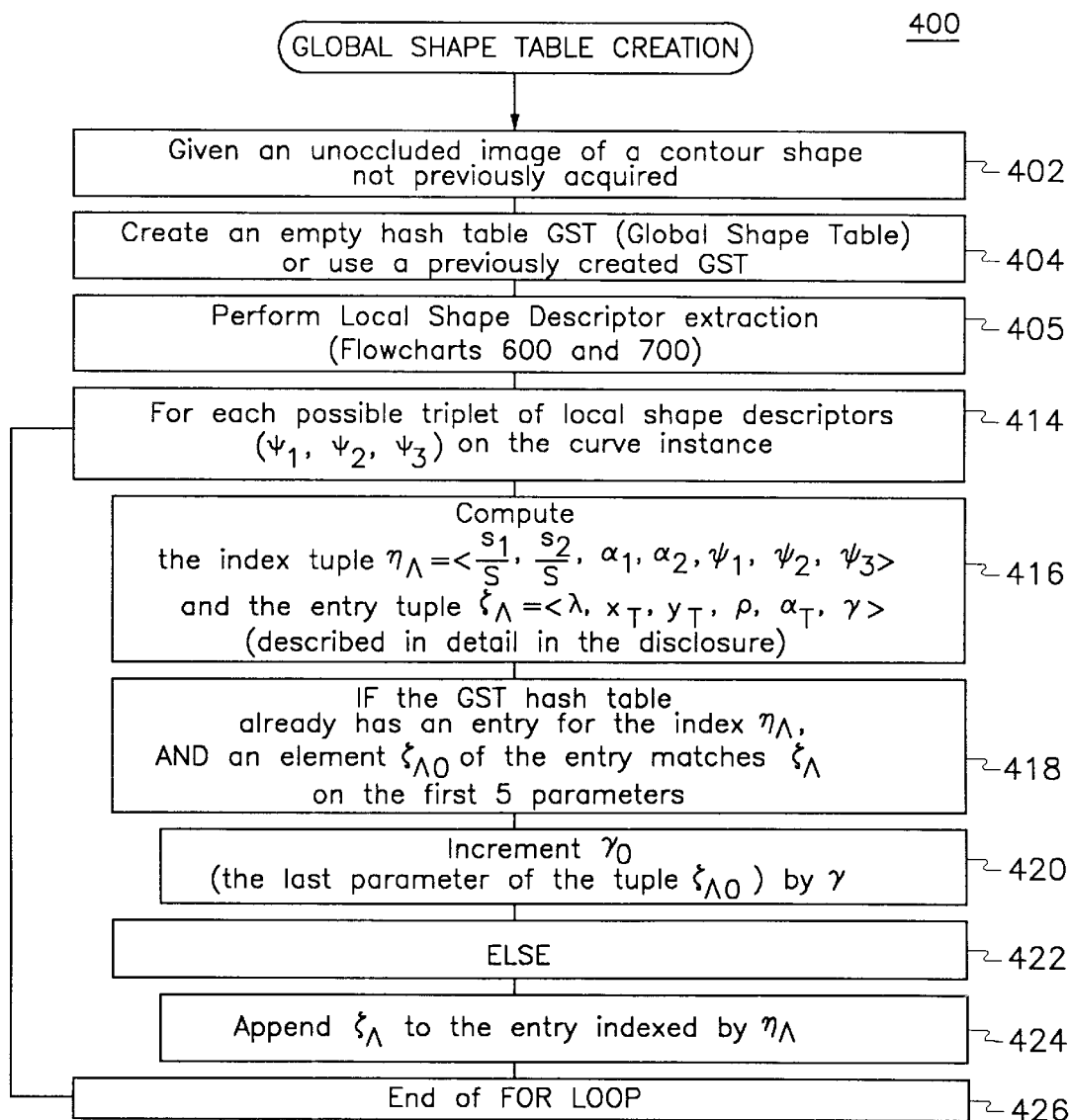
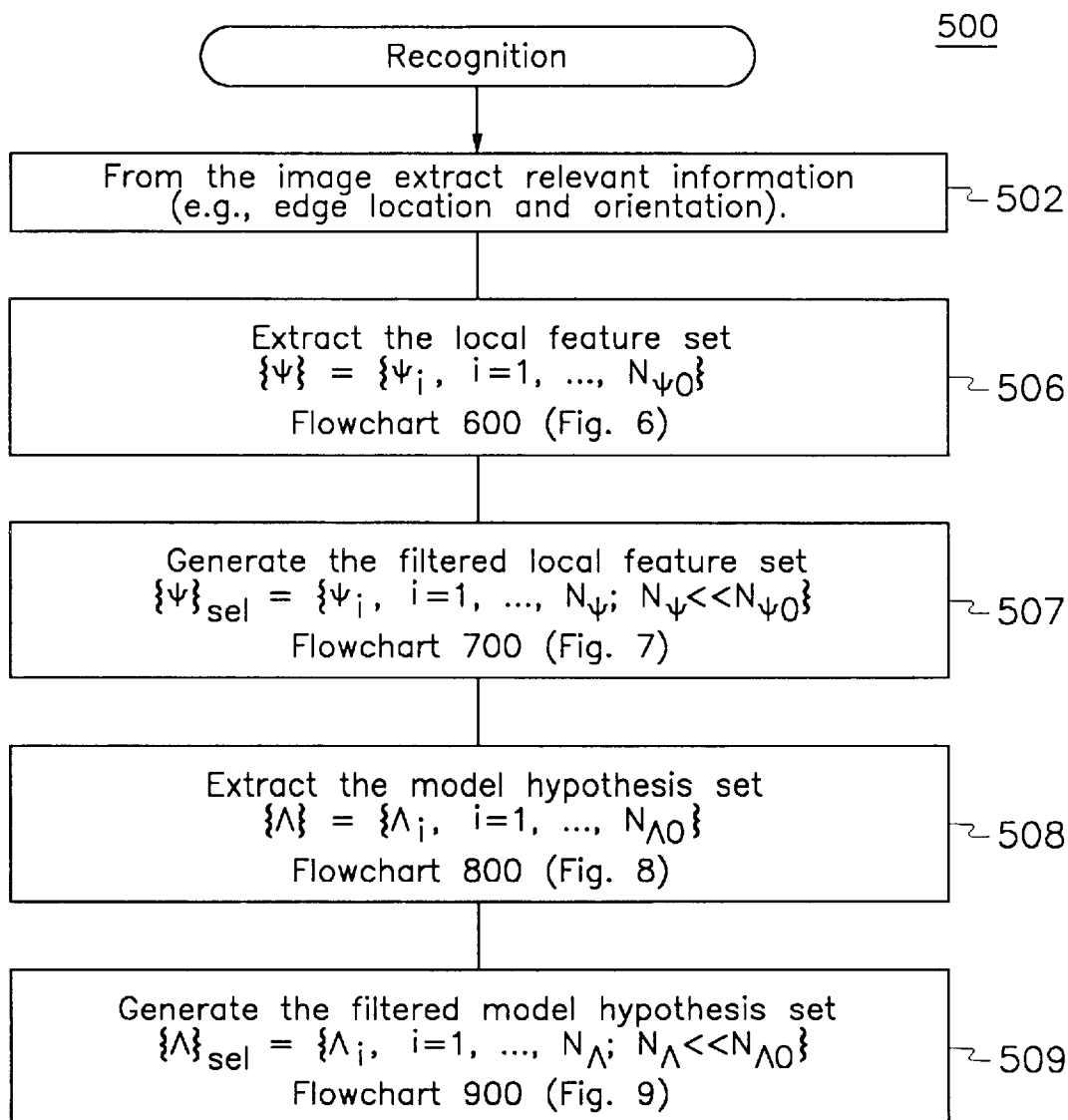


Fig. 4

*Fig. 5*

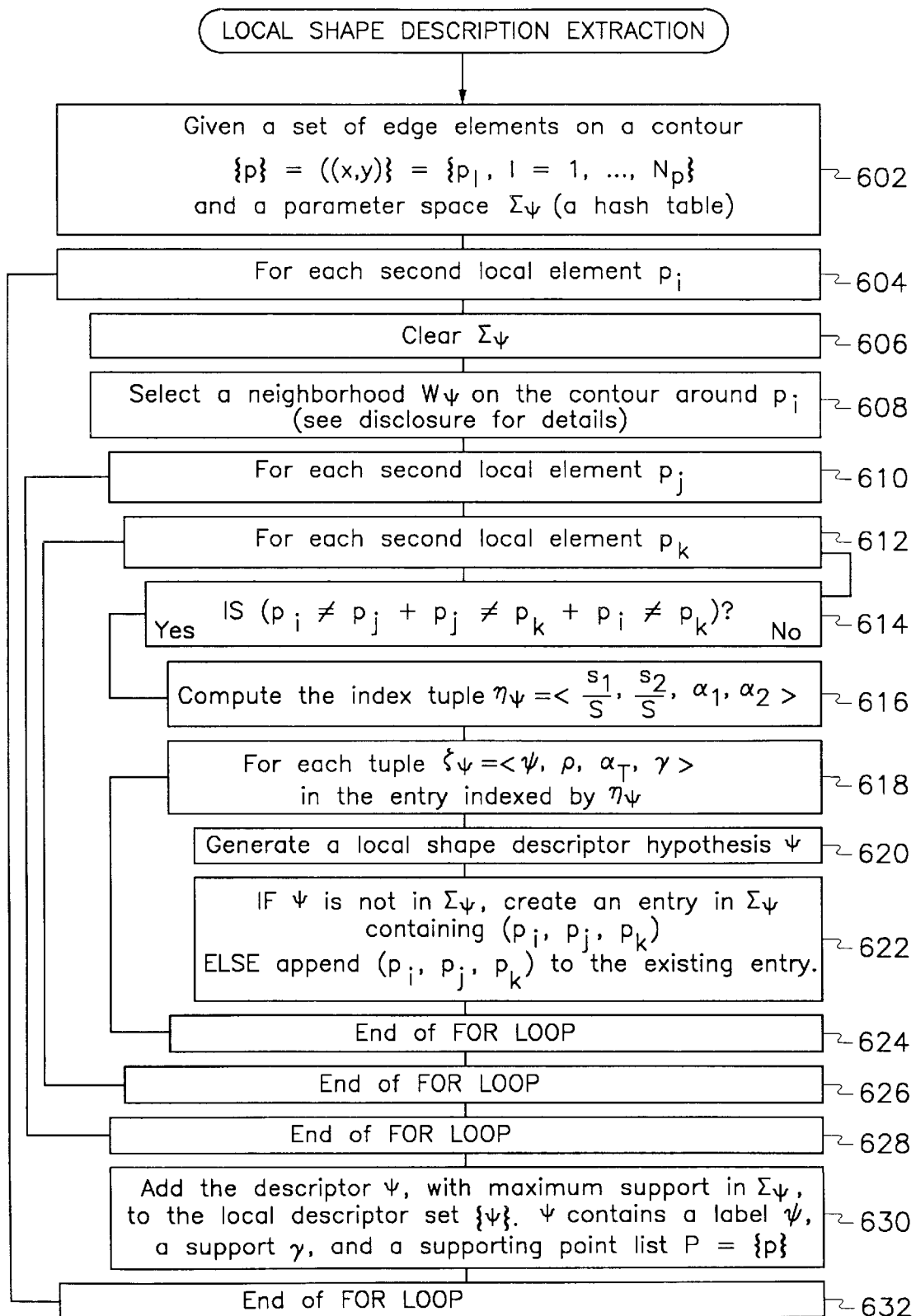
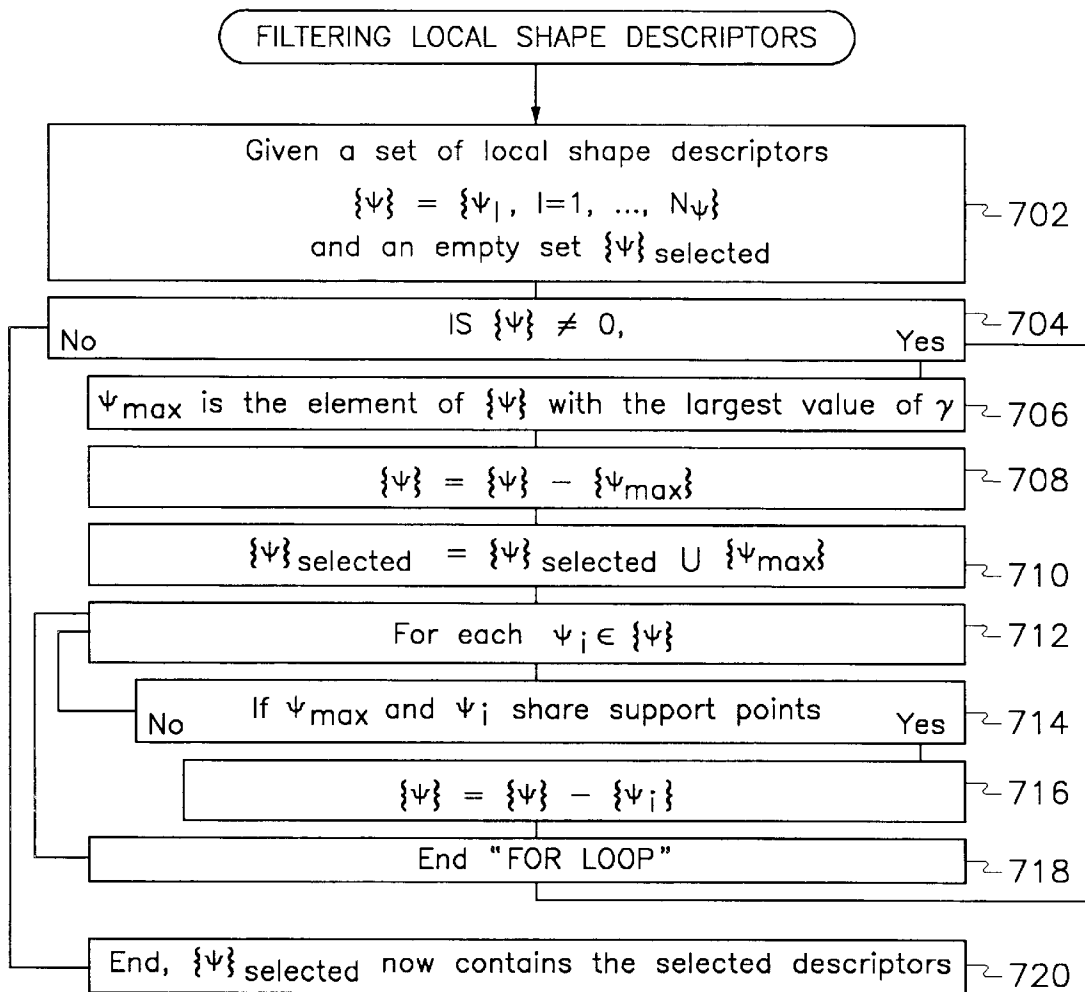


Fig. 6

*Fig. 7*

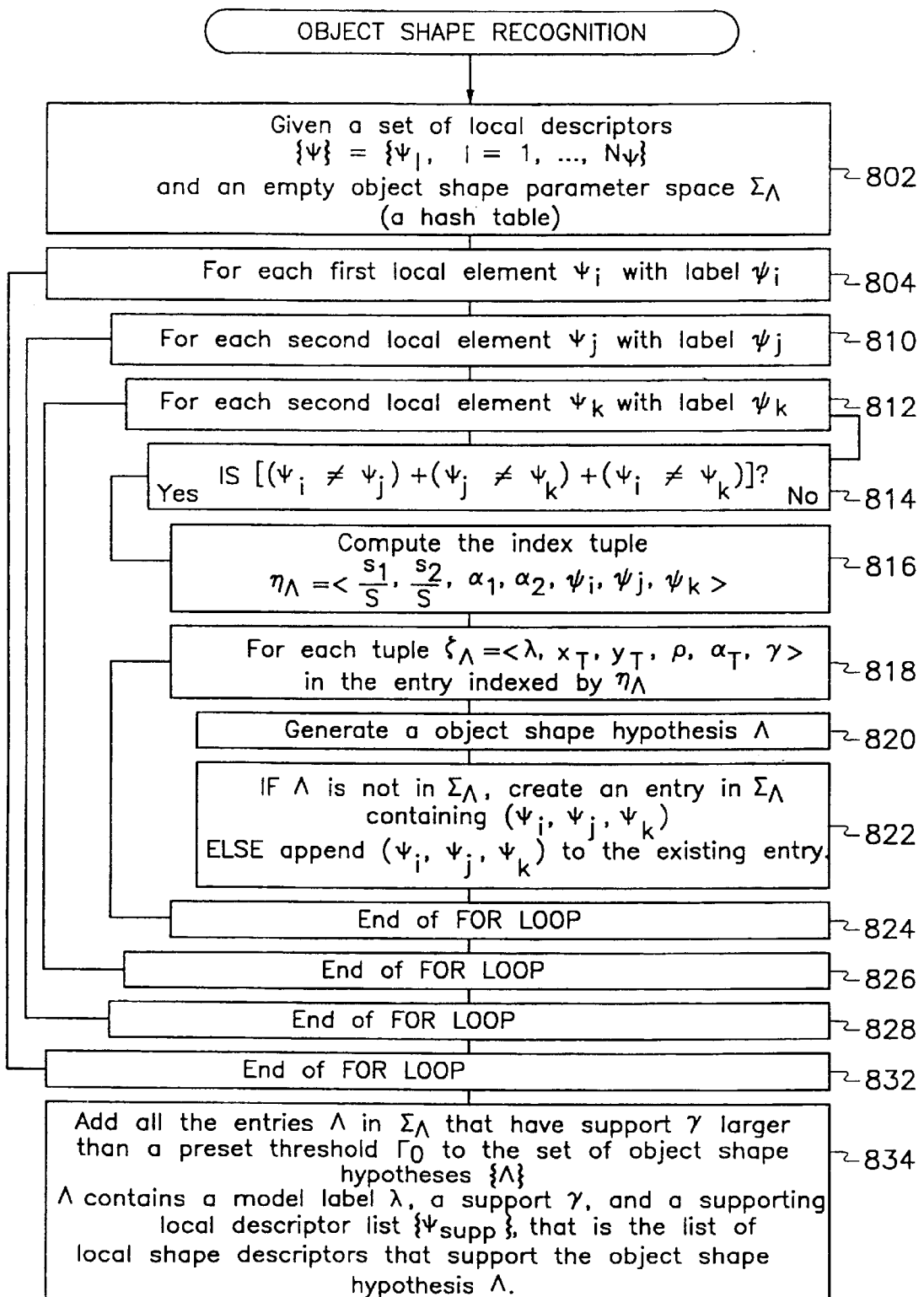


Fig. 8

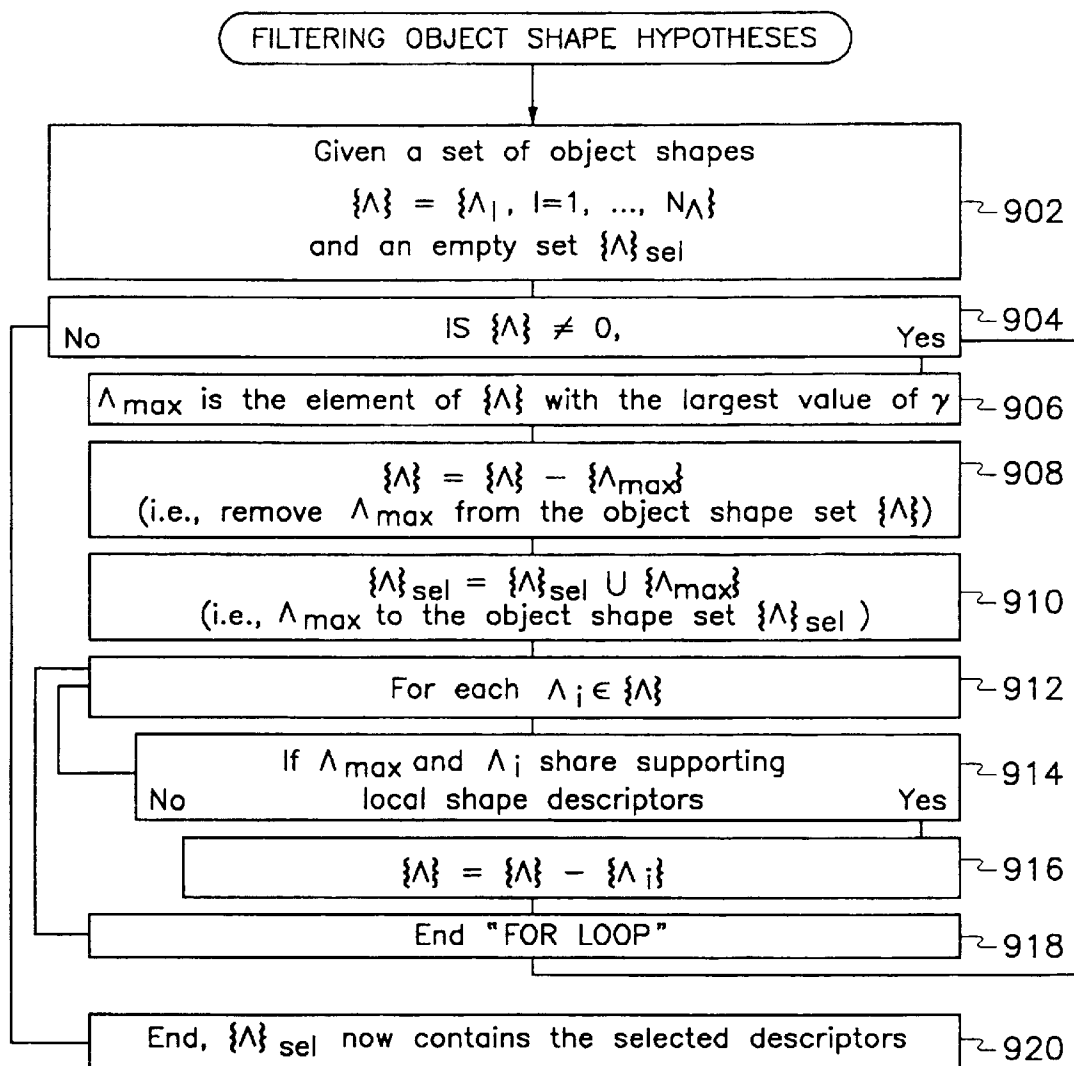


Fig. 9

GENERALIZED SHAPE AUTOCORRELATION FOR SHAPE ACQUISITION AND RECOGNITION

Matter enclosed in heavy brackets [] appears in the original patent but forms no part of this reissue specification; matter printed in italics indicates the additions made by reissue.

This application is a continuation of application Ser. No. 07/705,037, filed May 21, 1991, now abandoned.

TECHNICAL FIELD

The present invention relates generally to the recognition of complex 2D visual shapes, which could be derived from a viewing of 3D objects, and more particularly, to the automatic acquisition and recognition of complex visual shapes which are not easily modelled by conventional analytic techniques.

BACKGROUND ART

Automated systems for recognizing objects are increasingly being used in a large variety of technical fields, for example, biomedicine, cartography, metallurgy, industrial automation, and robotics. Moreover, many businesses are investing huge amounts of money and capital on the research and development of machine vision systems and related data processing systems which can automatically and accurately identify objects, often referred to as "target objects."

Automated object recognition systems are becoming more and more sophisticated as data processing techniques advance in sophistication. Most of the practical recognition systems in the conventional art employ methods involving the derivation of models to be used for recognizing objects in an image scene. An image scene in the context of this document is a two-dimensional (2D) representation, which for instance, could be derived from appearance data retrieved from a three-dimensional (3D) object at different viewpoints.

Furthermore, in most systems, data for deriving the models is inputted manually. However, in a few high-end automated systems, data can be learned via some sort of image capturing device, for example, a camera or scanner.

The model-based techniques have been conceptualized as having two phases: an object acquisition phase and a subsequent object recognition phase. More specifically, models of target objects are initially precompiled and stored during the acquisition phase, independently of the image scene. Then, the occurrences of these objects within an image scene are determined during the recognition phase by comparison of the sampled data to the stored models.

The task of recognizing objects in a scene is often complicated by rotation (vantage point), translation (placement), or scaling (size) of the object in a scene. In addition, the task may further be complicated by the partial concealment, or "occlusion", of a target object possibly caused by overlaps from other objects or some other adverse condition.

Some recognition systems employ "parametric" techniques, or mathematical parameter transforms. In parametric techniques, the spatial representation of an image in orthogonal coordinates is transformed into a representation based upon another coordinate system. Analysis of the image then takes place based upon the latter coordinate system. The methodology of using parametric techniques

originated in U.S. Pat. No. 3,069,654 to Hough, involving the study of subatomic particles passing through a viewing field.

Many of the commonly used parametric techniques are interrelated and have evolved over years of experimentation since the Hough patent. For a general discussion in regard to the use of parametric techniques for shape identification, see D. H. Ballard, "Parameter nets: A theory of low level vision," Proceedings of the 7th International Joint Conference on Artificial Intelligence, pp. 1068-1078, August 1981.

Well known conventional parametric techniques include "alignment techniques", "Hough Transform techniques", and "geometric hashing". Although not particularly relevant to the present invention, alignment techniques are discussed in the following articles: D. P. Huttenlocher, S. Ullman, "Three-Dimensional Model Matching from an Unconstrained Viewpoint," Proceedings of the 1'st International Conference on Computer Vision, pp. 102-111, London, 1987, and D. P. Huttenlocher, S. Ullman, "Recognizing Solid Objects by Alignment," Proceedings of the DARPA Image Understanding Workshop, vol. II, pp. 1114-1122, Cambridge, Massachusetts, April 1988. In regard to Hough Transform techniques, which also are not particularly relevant to the present invention, see D. H. Ballard, "Generalizing the Hough Transform to Detect Arbitrary Shapes," Pattern Recognition, vol. 13(2), pp. 111-122, 1981.

Geometric hashing is an often favored technique and is considered proper background for the present invention. In geometric hashing, models of objects are represented by "interest points." An orthogonal coordinate system is defined based on an ordered pair of interest points, sometimes referred to as the "basis pair." For example, the first and second interest points could be identified respectively as ordered pairs (0,0) and (1,0). Next, all other interest points are represented by their coordinates in the coordinate system.

The foregoing representation allows for comparison of objects which have been rotated, translated, or scaled, to the interest points of the model. Furthermore, the representation permits reliable comparison of the model to occluded objects, because the point coordinates of the occluded object in the sampled scene have a partial overlap with the coordinates of the stored model, provided both the model and scene are represented in a coordinate system derived from the same basis pair. However, occlusion of one or more of the basis points will preclude recognition of the object.

To avoid such a condition, interest points are represented in all possible orthogonal coordinate systems which can be derived from all of the possible basis pairs of interest points. Each coordinate is used to identify an entry to a "hash table." In the hash table, a "record" is stored which comprises the particular basis pair along with an identification of the particular model at issue.

During the object recognition phase, interest points initially are extracted from a scene. An arbitrary ordered pair is selected and used as the first basis pair. The coordinates of all other interest points in the scene are computed utilizing this basis pair. Each computed coordinate is compared to the coordinate entries of the hash table. If a computed coordinate and respective record (model, basis pair) appears in the hash table, then a "vote" is accorded the model and the basis pair as corresponding to the ones in the scene. The votes are accumulated in a "bucket." When a certain record (model basis pair) gets a large number of votes, then the record, and corresponding model, is adopted for further analysis.

Using the record, the edges of the specified model are compared against the edges in the scene. If the edges

correspond, then the object is considered matched to the model specified in the adopted record. If the edges do not match, then the current basis pair is discarded and a new basis pair is considered.

For further discussions in regard to geometric hashing, consider: Y. Lamdan, H. J. Wolfson, "Geometric hashing: a general and efficient model-based recognition scheme," Proceedings of the 2nd International Conference on Computer Vision, December 1988; J. Hong, H. J. Wolfson, "An Improved Model-Based Matching Method using Footprints," Proceedings of the International Conference on Pattern Recognition, Rome, Italy, November 1988; Y. Lamdan, J. T. Schwartz, H. J. Wolfson, "On recognition of 3-D objects from 2-D images," Proceedings of the IEEE International Conference on Robotics and Automation, vol. 3, pp.1, 1407-1413, Philadelphia, April 1988; and, finally, A. Kalvin, E. Schonberg, J. T. Schwartz, M. Sharir, "Two Dimensional Model Based Boundary Matching Using Footprints," The International Journal of Robotics Research, vol. 5(4), pp. 38-55, 1986.

Although to some degree effective, conventional parametric techniques, and specifically, conventional geometric hashing for the identification of complex visual shapes remains burdensome and undesirably time consuming in many circumstances. Moreover, geometric hashing is often unreliable because the performance of geometric hashing degrades significantly with only very limited amounts of clutter or perturbation in the sampled data. Geometric hashing can also be unreliable due to extreme sensitivity to quantization parameters. Finally, geometric hashing has limited index/model selectivity, oftentimes improperly accumulates excessive votes in each vote bucket, and has a limited number of useful buckets available in the hash tables.

DISCLOSURE OF INVENTION

The present invention provides for the acquisition and recognition of complex visual shapes.

During an acquisition phase, interest points are acquired from a target shape. A set of high dimensional indices are then derived from the interest points. Each of the indices has seven dimensions or more. The indices describe collectively a set of interest points along with the related geometry. For example, the indices can encode translation, rotation, and scale information about the shape(s) to be recognized.

During a recognition phase, interest points are observed from an inputted image. Further, it is recognized whether the target shape resides in the image by considering the high dimension indices.

In a specific implementation of the present invention, during acquisition and recognition, triplets of interest points of a target shape are transformed into local shape descriptors, which identify the interest points along with rotation and scale information about the triplet. Then, other triplets of the local shape descriptors are formed and are converted into the high dimension indices, which identify the local shape descriptors and also encode the translation, rotation, and scale information about the shape(s) to be recognized.

FEATURES AND ADVANTAGES OF THE INVENTION

The present invention overcomes the deficiencies of the prior art, as noted above, and further provides for the following additional features and advantages.

Generally, the present invention provides for the time-efficient automatic acquisition and recognition of complex visual shapes, which are difficult to identify using conventional parametric techniques, including conventional geometric hashing. In other words, the present invention enhances both accuracy and speed, which have traditionally been quid pro quos.

The global indices corresponding to a coordinate of an interest point are highly dimensional (7 or more characteristic dimensions). Moreover, these indices are very selective and invariant of Euclidean transformations. The highly dimensional indices overcome problems in conventional geometric hashing, such as limited index/model selectivity, excessive accumulation of votes in each bucket, limited number of useful buckets in the hash tables, and the undesirable extreme sensitivity to quantization parameters.

A model data base is provided which exhibits the properties of robustness (stability with respect to data perturbations), generalization, and recall from partial descriptions.

The methodology of the present invention can be implemented on commercial and inexpensive conventional computers.

The present invention exhibits both supervised and unsupervised learning of perceptual concepts, such as symmetry and the hierarchical organization of models into subparts.

It should be noted the above list of advantages is not exhaustive. Further advantages of the present invention will become apparent to one skilled in the art upon examination of the following drawings and the detailed description. It is intended that any additional advantages be incorporated herein.

BRIEF DESCRIPTION OF THE DRAWINGS

The present invention, as defined in the claims, can be better understood with reference to the text and to the following drawings.

FIG. 1 graphically illustrates at a high level the system architecture and methodology of the present invention wherein a local shape table and a global shape table are used to implement a parametric technique;

FIG. 2 shows a set of points from which triplets are formed wherein said points are either interests points from an image or local shape descriptors;

FIG. 3 illustrates a Flowchart for the initialization phase of the present invention wherein the local shape table of FIG. 1 is created;

FIG. 4 shows a Flowchart for the acquisition phase of the present invention wherein the global shape table of FIG. 1 is created;

FIG. 5 illustrates a high level Flowchart for the recognition phase of the present invention wherein target shapes in an image are recognized based upon those shapes that were acquired as models in FIG. 4;

FIG. 6 shows a low level Flowchart for the extraction of local shape descriptors during the recognition phase of the present invention;

FIG. 7 illustrates a low level Flowchart for the filtering of the local shape descriptors of FIG. 6 during the recognition phase;

FIG. 8 shows a low level Flowchart for the recognition of object shapes based upon an analysis of the filtered local shape descriptors of FIG. 7 during the recognition phase; and

FIG. 9 illustrates a low level Flowchart for the filtering of the object shapes of FIG. 8 during the recognition phase.

BEST MODE OF CARRYING OUT THE INVENTION

Table Of Contents

- A. Overview
- B. Initialization—Local Table Shape Creation
- C. Acquisition—Global Table Shape Creation
- D. Recognition
- 1. Local Shape Description Extraction
- 2. Object Shape Recognition
- A. Overview

The present invention presents a parametric technique, i.e., using parameter transforms for the initial acquisition and then subsequent recognition of complex arbitrary shapes in an image. The present invention envisions three phases for operation: (1) a one-time initialization phase, (2) an acquisition phase, and (3) a recognition phase. An overview of the phases is described below with reference to FIG. 1.

During the initialization phase, a local shape table **104** of FIG. 1 is created by inputting local shapes to be recognized during the recognition phase. After creation, the table **104** thereafter serves as a lookup table for local shapes during the recognition phase.

The local shape table **104** is structured to have an index η_{Ψ} for addressing a set of entries $\{\zeta_{\Psi}\}$. Each index η_{Ψ} and each entry ζ_{Ψ} are in the form of tuples. A "tuple" refers to a listing of elements or variables having a specific ordering. Tuples are denoted by the brackets "<...>" in the document.

The index η_{Ψ} and entry ζ_{Ψ} are derived by analyzing a "K-plet", a collection of K interest points, which could exist in an image, along with their associated local geometric properties, such as rotation and scale. In the preferred embodiment, "triplets" (K=3) of interest points are analyzed. During the initialization phase, the interest points are derived from synthetically generated images.

During the acquisition phase, target shapes in an image are learned and are used to derive the global shape table **108** of FIG. 1. First, local shapes are detected in the image. These local shapes are then used to acquire the target shape.

Specifically, interest points are initially acquired from target shapes in an inputted image, which target shapes are to be later recognized in any scene during the recognition phase. If the shape to be learned (acquired) is 2D, then as shown in FIG. 1, a single unoccluded image **102** of the target shape is presented to the system for acquisition.

The methodology as discussed in relation to 2D target shapes is equally applicable to 3D target objects. However, in the case of 3D target objects, different views are presented to the system in order to take into account a depth dimension, as is well known in the art. The views correspond to either discrete samplings of the Gaussian viewing sphere or different aspects of the model. For a discussion of modelling 3D objects, see J. Koenderink A. van Doorn, "The internal representation of solid shape with respect to vision," Biological Cybernetics, vol 32, pp. 211–216, 1979.

The interest points acquired from the image are analyzed in groups of K-plets. In the preferred embodiment, "triplets" (i.e., K=3) of interest points are analyzed. K-plets are used to derive local shape descriptors (LSD) $\{\Psi\}$. Each LSD is associated with a specific interest point (x,y) in the image **102**.

Further, the LSDs are filtered in order to choose the best ones by conventional means. In the preferred embodiment,

the conventional "best first" technique is utilized. The conventional "constraint satisfaction technique" could also be used to filter the LSDs. For a discussion of the constraint satisfaction technique, see D. H. Ballard, "Parameter nets: A theory of low level vision," Proceedings of the 7th International Joint Conference On Artificial Intelligence, pp. 1068–1078, August 1981; A. Califano, R. M. Bolle, and R. W. Taylor, "Generalized neighborhoods: A new approach to complex feature extraction," Proceedings of the IEEE Conference on computer vision and Recognition, June 1989; and R. Mohan, "Constraints Satisfaction networks for computer vision," Progress in Neural Networks, O. Omidvar Ed., Ablex, 1991.

Next, the filtered local shape descriptors are analyzed in groups of K-plets to derive the global shape table **108** of FIG. 1. In the preferred embodiment, "triplets" (K=3) of LSDs are analyzed to compute indicies η_{Λ} and entries ζ_{Λ} . The global shape table **108** is structured to have indices η_{Λ} for addressing a set of entries $\{\zeta_{\Lambda}\}$. Each of the indices η_{Λ} and entries ζ_{Λ} are in the form of tuples. Each tuple entry ζ_{Λ} comprises (1) a symbolic label λ , which uniquely identifies the model, (2) a set of geometric parameters β , used for computing the model's position, orientation, and scale in a scene, and (3) a support mechanism γ for keeping track of the votes for the same values of λ , β , and η_{Λ} . The geometric parameters β are invariant with respect to translation, rotation and scale transformations in subsequent scenes.

During the recognition phase, an image is generated from a scene and inputted into the system. Interest points are derived from the image. Moreover, local shape descriptors are obtained just as they are in the acquisition phase, discussed previously.

By analyzing K-plets of the LSDs, a set of object shape hypotheses Λ are mapped and accumulated in an object shape parameter space Σ_{Λ} . In the preferred embodiment, "triplets" (K=3) of LSDs are analyzed. Specifically, for each entry addressed by the index ζ_{Λ} for a K-plet of LSDs, a new model hypothesis $\Lambda=\langle\lambda,\phi\rangle$ is formed. The model hypothesis includes the model label λ and its corresponding registration parameters ϕ , computed from geometric parameters β , which encode information relating to the geometry of the K-plet. The registration parameters ϕ uniquely identify the shape, rotation of the shape, and the shape's scale.

An object shape hypothesis defined by the model label λ and registration parameters ϕ is added to the global shape parameter space $\Sigma_{\Lambda}(\lambda, \phi)$.

Finally, the object shape hypotheses $\{\Lambda\}$ are filtered in the parameter space $\Sigma_{\Lambda}(\lambda, \phi)$ to select the best, mutually compatible, object shapes $\{\lambda\}$. The set of filtered object shape hypotheses $\{\lambda\}_{selected}$ essentially comprise the shapes recognized and located in the scene.

In a general sense, the present invention is performing "spatial autocorrelation." "Spatial autocorrelation" in the context of this disclosure means that non-local spatial information is being correlated. In other words, groupings of interest points or LSDs, called K-plets which are spread across an image, are correlated to derive local shape descriptors Ψ , or object shapes Λ , which embed global shape information.

It should be noted that spatial autocorrelation to some extent involves the conventional Hough transform technique. However, in the Hough transform technique, only a single interest point is considered at a time, unlike in the present invention where a K-plet is considered. For a further discussion regarding spatial autocorrelation, see A. Califano, "Feature recognition using correlated information contained in multiple neighborhoods," Proceedings of the 7th National

Conference on Artificial Intelligence, July 1988, pp. 831–836, and also, A. Califano, R. M. Bolle, and R. W. Taylor, “Generalized neighborhoods: A new approach to complex feature extraction,” Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, June 1989.

B. Initialization—Local Table Creation

During the initialization phase, the local shape table **104** of FIG. 1 is created. The methodology for creating the foregoing table is shown and will be described below in regard to FIG. 3.

The creation of the local shape table need only be performed once, and then the results can be stored in a transportable file (perhaps, for example, on a magnetic medium). The creation of the local shape table **104** can be considered an initialization process. In other words, when new target shapes are added to the system, the local shape table need not be recreated.

As illustrated in FIG. 3, the local shape table is created based upon the local shapes which are sought to be recognized by the system. First, an empty local shape table is created. The local shape table is created in the form of a lookup table, where an index η corresponds with a unique set of entries $\{\zeta\}$.

In a FOR LOOP enclosed by blocks **306** and **330**, data corresponding to each of the curves desired to be recognized are inputted to the table **104**. In this document, these curves are called “local shapes” and are accorded a specific local shape descriptor label ψ for identification purposes.

The local shapes could be, for example, lines, circular arcs, elliptic arcs (minima and maxima of curvature), and so forth. Large sets of local shapes can be considered. In this regard see A. P. Blicher, “A shape representation based on geometric topology: Bumps, Gaussian curvature, and the topological zodiac,” Proceedings of the 10th International Conference on Artificial Intelligence, pp. 767–770, August 1987.

These local shapes are “synthetically” drawn for the system. “Synthetically” means that the local shapes are automatically generated by the system. In other words, the shapes need not be created or drawn on a display screen or some other visual device.

Once local shapes are synthetically drawn, the manner in which they are quantized, or digitized, depends upon the way the local shapes are scaled or oriented. Accordingly, via a FOR LOOP enclosed by blocks **308** and **328**, the local shapes are drawn at a few different scales and orientations (translations and rotations) to enhance the integrity of the ultimate models.

Next, each local shape is synthetically positioned symmetrically on an $M \times M$ array, as indicated in a block **310**. At a block **312**, the center point p_1 of the local shape is identified. Two other points p_2 and p_3 are arbitrarily selected along the local shape, to thereby derive a triplet of interest points. Each triplet forms a specific triangle, having the three points p_1 , p_2 , and p_3 , and three sides s_1 , s_2 , and s_3 . The local shape descriptor labels ψ_i are associated with each of the points p_i .

A FOR LOOP enclosed by blocks **314** and **326** operates upon each triplet of interest points p_1 , p_2 , p_3 in order to derive the indices and entries in the local shape table **104**.

At a block **316**, for each triplet, an index η_ψ and a corresponding entry ζ_ψ is computed and ζ_ψ is entered into the table **104** in the form of tuple at the location addressed by the index η_ψ .

Essentially, the index η is an address in the lookup table. The index η_ψ is generated as follows:

$$\eta_\psi = \langle \frac{s_1}{S}, \frac{s_2}{S}, \alpha_1, \alpha_2 \rangle$$

The two ratios s_1/S and s_2/S , where $S=s_1+s_2+s_3$, essentially define the geometry of the triangle, as shown in FIG. 2. As further indicated in FIG. 2, the two angles α_1 and α_2 describe the 1st order properties of the local shape with respect to the given triplet.

For the first point p_1 , any properties higher than 0th order are not used, because these properties are more noise sensitive and could bias the indices for all possible triangles generated for this feature.

The tuple entry ζ , corresponding to the index η , is computed as follows:

$$\zeta_\psi = \langle \psi, \rho, \alpha_T, \gamma \rangle$$

for each triplet, the tuple $\langle \psi, \rho, \alpha_T, \gamma \rangle$ is inserted into the local shape table. Here, ψ is the symbolic label for the local shape. As shown in FIG. 2, α_T is the angle between the vector t_{23} and the normal n to the contour at p_1 required to recover the shape orientation from the triplet. $\rho=\mu/S$, where μ is a constant, is used for scale normalization.

After each index η and corresponding set of entries ζ are computed for each triplet of interest points, as shown in a block **318**, the foregoing tuples are compared with other indices and entries in the local shape table **104** to determine if it is redundant. First, it is determined whether the table **104** already has entries ζ_ψ at index η_ψ .

If entries exist, then the parameters ψ , ρ , α_T of the entry ζ_ψ are compared with the like parameters in the already-existing entries at η_ψ . If these parameters match, then the support mechanism γ is accorded a vote, as indicated in a block **320**.

However, if the table **104** does not have a matching entry at index η_ψ , then the entry ζ_ψ is entered at the location addressed by index η_ψ in the table **104**, as shown in a block **324**.

After the FOR LOOPS **314,326; 308,328; and 306,330** have completed operation, the local shape table has been completed.

C. Acquisition—Global Table Creation

During the acquisition phase of the present invention, target shapes are stored in the system for use as models to later recognize such shapes in a scene.

FIG. 4 illustrates the methodology for constructing the global shape table **108** of FIG. 1. The global shape table **108** is created when the images to be learned, or acquired, by the system are inputted. Note that the methodology of FIG. 4 is similar to that of FIG. 3. Moreover, similar steps within referenced blocks are denoted with similar reference numerals in the FIGS. 3 and 4 for easy comparison.

Initially, as indicated in a block **402** of FIG. 4, an unoccluded image is inputted into the system. The image may be a photograph, an image generated from a camera, or the like. With reference to FIG. 2, the image is inputted so that the position of the target shape's center of mass (x_0, y_0) is computed. A scale (e.g., perimeter) μ and orientation vector d are then assigned. Further, a label λ_λ is provided to identify the target shape.

At block **404**, an empty global shape table is created or an existing global shape table is utilized. Similar to the local shape table, the global shape table is created in the form of a lookup table, where an index η corresponds with a unique set of entries $\{\zeta\}$.

Next, at a block **405**, the local shapes of the image are detected. As an example, a local shape **112** is shown in FIG. 1. Essentially, the edges of the image along with the locations of the edges are considered and local shapes are computed. Moreover, the location of the local shapes is recorded. Finally, from the foregoing information, local shape descriptors $\Psi_1-\Psi_n$ are derived, as indicated at reference numeral **106** of FIG. 1. The procedure for detecting local shapes will be more fully discussed in regard to FIGS. **5** and **6** hereinafter.

For each triplet of local shape descriptors Ψ_1, Ψ_2 , and Ψ_3 , a tuple index η_Λ and a corresponding tuple entry ζ_Λ is derived for the global shape table **108** of FIG. 1. This procedure is implemented via the FOR LOOP enclosed by blocks **414** and **426**, which is analogous to the FOR LOOP enclosed by blocks **314** and **326** in FIG. 3. However the index η_Λ of FIG. 4 has more parameters than the index η_Ψ of FIG. 3.

In FIG. 4, the index η_Λ is computed and is specified as follows:

$$\eta_\Lambda = \langle \frac{s_1}{S}, \frac{s_2}{S}, \alpha_1, \alpha_2, \psi_1, \psi_2, \psi_3 \rangle$$

In comparison to the index η_Ψ , the index η_Λ comprises three additional parameters, namely, ψ_1, ψ_2, ψ_3 , which are essentially the identification of the LSD triplet. Further, as with the triplets of interest points used to derive the local shape descriptors, the LSD triplets form a triangle with sides s_1, s_2, s_3 , as shown in FIG. 2. The two ratios s_1/S and s_2/S , where $S=s_1+s_2+s_3$, define the geometry of the triangle, as shown in FIG. 2. As further indicated in FIG. 2, the two angles α_1 and α_2 describe the 1st order properties of the local shape with respect to the given triplet.

Again, for the first LSD Ψ_1 at point p_1 , any properties higher than 0th order are not used, because these properties are more noise sensitive and could bias the indices for all possible LSD triangles generated for this feature.

The tuple entry ζ_Λ , corresponding to the index η_Λ , is computed as follows:

$$\zeta_\Lambda = \langle \lambda, x_T, y_T, \rho, \alpha_T, \gamma \rangle$$

For each LSD triplet, the tuple $\langle \lambda, x_T, y_T, \rho, \alpha_T, \gamma \rangle$ is inserted into the global shape table **108** at the location indexed by the index η_Λ .

As shown in FIG. 2, the position (x_T, y_T) is the center of the shape's area in the new right hand coordinate system defined by the normalized vector t_{23} and its corresponding orthonormal counterpart d , the ratio $\rho=\lambda/S$, and the angle α_T between the two vectors t_{23} and d . It should be noted that the geometric parameters $\beta=((x_T, y_T), \rho, \alpha_T)$ are also scale and rotation invariant, and are used to recover position, orientation and scale of a feature from the corresponding **3** local shape descriptors $\{\psi_i, \psi_j, \psi_k\}$.

After each index η_Λ and corresponding entry ζ_Λ are computed for each LSD triplet, as shown in a block **418**, the foregoing tuples are compared with other indices and entries in the global shape table **108** to determine if it is redundant. First, it is determined whether the table **108** already has an index η_Λ .

If one exists, then the parameters $\lambda, x_T, y_T, \rho, \alpha_T$ of the entry ζ_Λ are compared with the like parameters in the already-existing entry. If these parameters match, then the support mechanism γ (initially set to one) of the matching entry ζ_Λ is accorded a vote, as indicated in a block **420**.

However, if the table **108** does not have a matching entry at index η_Λ , then the entry ζ_Λ is entered at the location addressed by index η_Λ in the table **108**, as shown in a block **424**.

After the FOR LOOP **414,426** has completed operation, the global shape table has been completed. Accordingly, the system can now recognize shapes in target images.

D. Recognition

The methodology envisioned by the present invention for recognition of target shapes and objects will be discussed in detail hereafter in regard to FIGS. **5-9**. FIG. **5** illustrates a high level Flowchart **500** showing the overall recognition methodology. FIGS. **6-9** comprise low level Flowcharts **600-900** describing in detail respective blocks **506-509** of FIG. **5**.

With reference to block **502** of FIG. **5**, the edges of an image to be recognized are extracted from a scene. This procedure is well known in the art.

At the next block **506**, a "local feature set" is derived. The set comprises the LSDs $\Psi_1-\Psi_N$ for local shapes of the image. For example, a local shape is shown as reference numeral **112** in FIG. 1. A detailed description of this procedure will be discussed with reference to Flowchart **6**, entitled "Local Shape Description Extraction" of FIG. **6**, hereinafter.

Next, at the block **507**, the local feature set $\{\Psi_1-\Psi_N\}$ generated in the previous step is filtered using any conventional filtering technique. The resultant filtered set of LSDs is extremely smaller than the original set of LSDs $\Psi_1-\Psi_N$. The resultant filtered set is illustrated at reference numeral **106** in FIG. 1.

In the preferred embodiment, the best first technique is utilized to filter the LSDs, which technique is well known in the art. As mentioned previously, any conventional filtering technique could be utilized, such as the constraint satisfaction technique. The specific implementation of this technique is shown in Flowchart **700**, entitled "Filtering Local Shape Descriptors" of FIG. 7. Due to the well known nature and applicability of this technique, no discussion of the Flowchart **700** is undertaken.

At the next block **508**, a set of model hypotheses $\Lambda_1-\Lambda_N$ is derived. This procedure is set forth in Flowchart **800**, entitled "Object Shape Recognition", of FIG. 8.

The model hypothesis set $\{\Lambda_1-\Lambda_N\}$ generated in the previous step is filtered using any conventional filtering technique. The resultant filtered set is extremely smaller than the original set.

In the preferred embodiment, the first best technique is again utilized for filtering. The specific implementation of this technique is shown in Flowchart **900**, entitled "Filtering Object Shape Hypotheses", of FIG. 9. Due to the well known nature and applicability of this technique, no discussion of the Flowchart **900** is undertaken.

The model hypotheses which survive the foregoing filtering, if any, are adopted as the shapes which are recognized in the scene. Thus, recognition has occurred.

1. Local Shape Description Extraction

FIG. **6** illustrates an exemplary Flowchart **600** for the extraction of local shapes during the recognition phase. A local feature set is derived in FIG. **6**. The set comprises the LSDs $\Psi_1-\Psi_N$ corresponding with local shapes of the image. For example, a local shape **112** is shown in FIG. 1.

First, at a block **602**, N interest points p of an image are extracted from a scene using any of the numerous conventional techniques. The location of each interest point p is identified by orthogonal coordinates (x,y) . This procedure is well known in the art. Moreover, a temporary parameter

space Σ_Ψ (a hash table) is generated for use in the proceeding steps of the Flowchart 600.

Next, a FOR LOOP enclosed by blocks 604 and 632 commences for each interest point p_i of $\{p\}$. In the FOR LOOP 604,632, the parameter space Σ_Ψ is first cleared, as indicated in block 606.

At block 608, a neighborhood W_Ψ on the contour around p_i is selected. As will be discussed further below, triplets of interest points will be selected from the neighborhood W_Ψ . Looked at another way, the neighborhood W_Ψ prevents the interest points in triplets from being undesirably scattered all over the image.

To derive the neighborhood W_Ψ , edges are linked into contours using simple "eight neighbors connectivity." Given a point $p_i=(s_i, y_i)$ on a contour, points p_1 are symmetrically sampled around p_i along the contour, thereby generating a neighborhood set $W=\{p_1=(x_1, y_1), -N \leq 1 \leq N\}$. The sampling step is proportional to the length of the longest symmetrical interval around the point p_i for which the tangent behavior is monotonic on a coarse scale, and the total tangent variation is less than $2\pi/3$. It is also inversely proportional to the total variation in tangent angle on the symmetrical interval if less than $2\pi/3$. An assumption is that faster variation in tangent correspond to smaller features which, accordingly, require finer sampling for detection. In this way, a quasi-scale invariant sampling of the contours is achieved. Next, triplets of interest points are derived. All possible combinations are formed of the point (x_i, y_i) in conjunction with two others p_j and p_k from the set $\{p_1=(x_1, y_1), -N \leq 1 \leq N\}$. These combinations are made in FOR LOOPS 610,628 (p_j) and 612,626 (p_k), as shown. At block 614, it is insured that the same point is not selected twice when constructing the triplet combinations.

At block 616, the index $\eta_\Psi=(s_1/S, s_2/S, \alpha_1, \alpha_2)$, described herein in the section entitled, "Local Shape Table Creation", is computed for each triplet.

At block 618, the tuple index η_Ψ is now used to index into the local shape table 104 in order to retrieve the corresponding tuple entry $\zeta_\Psi=<\lambda, \rho, \alpha_T, \gamma>$.

Next computations are performed to generate a local shape descriptor hypothesis Ψ , as indicated in block 620. Specifically, from the ρ and α_T of the tuple entry ζ_Ψ , the possible scalar orientations of the contour are created. Note that the location of the contour is at p_i , and accordingly, the location of the contour need not be computed as with global shapes. Thus, only scale and rotation information need be computed.

At block 622, the parameter space Σ_Ψ is examined to determine whether an LSD hypothesis Ψ already exists in the parameter space Σ_Ψ for the contour at issue. If the LSD hypothesis Ψ exists, then the existing LSD hypothesis Ψ is merely updated by appending points p_i, p_j, p_k to it. If the LSD hypothesis Ψ does not exist, then the entry ζ_Ψ for Ψ having the three points p_i, p_j, p_k is created in the parameter space Σ_Ψ .

After completion of FOR LOOPS 618,624; 612,626; and 610,628, an LSD hypothesis Ψ will have a high vote count, based upon votes accorded it by the points p_1, p_j, p_k . These points along with the corresponding LSD hypothesis Ψ belong to the particular contour at issue.

Thus, at the next block 630 the LSD hypothesis Ψ is adopted as the local shape descriptor Ψ for p_i . For each point p_i , there exists a local shape descriptor Ψ having a label ψ , a support mechanism γ , and a supporting point list $P=\{p\}$.

After completion of Flowchart 600, a large set of local shape descriptors $\{\Psi\}$ with exactly one LSD and corresponding to each interest point p_i in the image is available for further mathematical manipulation.

In order to winnow down the number of LSDs Ψ and to enhance the integrity of the LSDs, the local feature set $\{\Psi_1-\Psi_N\}$ is filtered using any conventional filtering technique as discussed previously. In the preferred embodiment, a conventional best first technique is utilized and is illustrated at Flowchart 700 of FIG. 7. The Flowchart 700 is self explanatory. The resultant filtered set of LSDs, $\{\Psi\}_{selected}$, is extremely smaller than the original set of LSDs $\Psi_1-\Psi_N$. The resultant filtered set is illustrated at reference numeral 106 in FIG. 1.

2. Object Shape Recognition

FIG. 8 shows a Flowchart 800 for recognizing an object based upon the local feature set $\{\Psi\}_{selected}$ comprising the filtered LSDs Ψ . Flowchart 800 proceeds very similarly with respect to Flowchart 600. Accordingly, similar steps within referenced blocks are denoted with similar reference numerals in the Flowcharts 800 and 600 for easy comparison. However, some differences in methodology do exist.

In Flowchart 800, the focus is upon local curve shapes, not upon interest points as in Flowchart 600. Further, triplets in Flowchart 800 are formed from LSDs Ψ , not with interest points as in Flowchart 600. Finally, the global shape table 108 is consulted during the methodology of Flowchart 800, not the local shape table 104 as in Flowchart 600. Note that the index η_Ψ corresponding with the global shape table 108 is larger than the index η_Λ corresponding with the local shape table 104.

With reference to Flowchart 800, at a block 802, the N filtered LSDs Ψ corresponding to the image are acquired.

Moreover, a temporary parameter space Σ_Λ (a hash table) is generated for use in the proceeding steps of the Flowchart 800.

Next, a FOR LOOP enclosed by blocks 804 and 832 commences for each LSD Ψ of the local feature set $\{\Psi\}_{selected}$. In the FOR LOOP 804,832, the parameter space Σ_Λ is first cleared, as indicated in block 806.

Next, triplets of LSDs Ψ are derived. All possible combinations are formed of the LSD Ψ_i in conjunction with two others Ψ_j and Ψ_k from the local feature set $\{\Psi\}_{selected}$. These combinations are made in FOR LOOPS 810,828 (Ψ_j) and 612,626 (Ψ_k), as shown. At block 814, it is insured that the same LSD is not selected twice when constructing the triplet combinations.

At block 816, the index $\eta_\Lambda=<s_1/S, s_2/S, \alpha_1, \alpha_2, \psi_i, \psi_j, \psi_k>$, described herein in the section entitled, "Global Shape Table Creation", is computed for each LSD triplet.

At block 818, the index η_Λ is now used to index into the global shape table 108 in order to retrieve the corresponding entry $\zeta_\Lambda=<\lambda, x_T, y_T, \rho, \alpha_T, \gamma>$.

Next, computations are performed to generate an object shape hypothesis Λ , as indicated in block 820. Specifically, from the x_T, y_T, ρ, α_T of the tuple entry ζ_Λ , the possible orientations of the contour are created. In other words, scale, rotation, and translation information is computed.

At block 822, the parameter space Σ_Λ is examined to determine whether an object shape hypothesis Λ already exists in the parameter space Σ_Λ . If the object shape hypothesis Λ already exists, then the existing object shape hypothesis Λ is merely updated by appending (Ψ_i, Ψ_j, Ψ_k) to it and incrementing its vote. If the object shape hypothesis Λ does not exist, then the object shape instance Λ having the triplet Ψ_i, Ψ_j, Ψ_k is created in the parameter space Σ_Λ .

After completion of FOR LOOPS 818,824; 812,826; and 810,828, a number of object shape hypotheses $\{\Lambda\}$ will have a high vote count, based upon votes accorded it by the LSDs Ψ_i, Ψ_j, Ψ_k .

Thus, at the next block 834, the object shape hypotheses $\{\Lambda\}$ having a votes γ above a preset threshold Γ_0 are

adopted. For each Λ , there exists a unique model label λ , a support γ , and a supporting LSD list $\{\Psi_{supp}\}$.

After completion of Flowchart 800, a set of object shape hypotheses $\{\Lambda\}$ corresponding with each LSD triplet in the image is available.

In order to enhance the integrity of the set of object shape hypotheses $\{\Lambda\}$, they are filtered using any conventional filtering technique as discussed previously. In the preferred embodiment, a conventional best first technique is utilized and is illustrated at Flowchart 900 of FIG. 9. The Flowchart 900 is self-explanatory. The resultant filtered set of LSDs, $\{\Lambda\}_{selected}$, is illustrated at reference numeral 110 in FIG. 1. The $\{\Lambda\}_{selected}$ constitute the object shape recognized and located in the image.

The foregoing description of the preferred embodiment of the present invention has been presented for purposes of illustration and description. It is not intended to be exhaustive or to limit the present invention to the precise form disclosed, and obviously many modifications and variations are possible in light of the above teachings. The particular embodiments were chosen and described in order to best explain the principles of the present invention and its practical application to those persons skilled in the art and to thereby enable those persons skilled in the art to best utilize the present invention in various embodiments and with various modifications as are suited to the particular use contemplated. It is intended that the scope of the present invention be broadly defined by the claims appended hereto.

The following is claimed:

1. A computer-implemented method for recognizing objects and for automatic acquisition of models of objects, comprising the steps of:

- (1) acquiring a model of an object, comprising,
 - (a) digitizing said object to generate a digitized image;
 - (b) detecting local shapes in said digitized image;
 - (c) grouping three or more [noncontinuous] *noncontiguous* combinations of said local shapes to generate a first index;
 - (d) generating an entry for each group of local shapes consisting of a name of said digitized image, and information about the translation, rotation, and scale of said digitized image; and
 - (e) storing said entry in a shape table at said first index; and
- (2) recognizing a target object that appears in a physical scene, comprising,
 - (a) digitizing said target object to generate a digitized target image;
 - (b) detecting [global] *local* shapes in said digitized target image;
 - (c) grouping three or more [noncontinuous] *noncontiguous* combinations of said [global] *local* shapes to generate a second index;
 - (d) accessing said shape table and retrieving entries from said shape table that correspond to said second index;
 - (e) collecting said retrieved entries from said shape table into a vote table; and
 - (f) selecting said retrieved entry with a highest vote in order to recognize said target image.

2. The method of claim 1, further comprising the steps of: creating a local lookup table;

entering local shapes to be recognized by the system into said local lookup table by synthetic generation; and using said local lookup table to detect local shapes in said digitized image.

3. The method of claim 1, further comprising the steps of: transforming three or more edge points of said digitized image into local groupings;

analyzing said local groupings in a parameter space; comparing the geometric characteristics of said local groupings with local shape entries in a local lookup table;

according votes to said local shape entries which correspond with said geometric characteristics; and adopting local shape descriptors corresponding to local shape entries which have a high number of votes.

4. The method of claim 1, further comprising the steps of: converting three or more local shapes of said digitized target image into global groupings;

analyzing said global groupings in a parameter space; comparing the geometric characteristics of said global groupings with global shape entries in said shape table; according votes to said global shape entries which correspond with said geometric characteristics; and adopting global shape descriptors corresponding to global shape entries which have a high number of votes.

5. A computer-based system for recognizing objects and for automatic acquisition of models of objects, comprising:

- (a) means for acquiring a model of an object, comprising,
 - (1) means for digitizing said object to generate a digitized image;
 - (2) means for detecting local shapes in said digitized image;
 - (3) means for grouping three or more [noncontinuous] *noncontiguous* combinations of said local shapes to generate a first index;
 - (4) means for generating an entry for each group of local shapes consisting of a name of said digitized image, and information about the translation, rotation, and scale of said digitized image; and
 - (5) means for storing said entry in said shape table at said first index; and
- (b) means for recognizing a target object that appears in a physical scene, comprising,
 - (1) means for digitizing said target object to generate a digitized target image;
 - (2) means for detecting local shapes in said digitized target image;
 - (3) means for grouping three or more [noncontinuous] *noncontiguous* combinations of said [local] shapes to generate a second index;
 - (4) means for accessing said shape table and for retrieving entries from said shape table that correspond to said second index;
 - (5) means for collecting said retrieved entries from said shape table into a vote table and for selecting said retrieved entry with a highest vote in order to recognize said target image.

6. The computer-based system of claim 5, further comprising:

- (a) a camera adapted for viewing said target object and producing a signal indicative of said target object; and
- (b) a digitizer, connected to said camera, for digitizing said target object represented by said signal and storing said digitized target image in a storage medium.

7. The computer-based system of claim 5, further comprising:

- a local shape table;
- means for entering local shapes to be recognized into said local shape table by synthetic generation; and

15

means for using said local shape table to detect local shapes in said digitized image.

8. The computer-based system of claim 5, further comprising:

means for transforming three or more edge points of said digitized image into local groupings;

means for analyzing said local groupings in a parameter space;

means for comparing the geometric characteristics of said local groupings with local shape entries in a local lookup table;

means for according votes to said local shape entries which correspond with said geometric characteristics; and

means for adopting local shape descriptors corresponding to local shape entries which have a high number of votes.

9. The computer-based system of claim 5, further comprising:

16

means for converting three or more local shapes of said digitized target image into global groupings;

means for analyzing said global groupings in a parameter space;

means for comparing the geometric characteristics of said global groupings with global shape entries in said global shape table;

means for according votes to said global shape entries which correspond with said geometric characteristics; and

means for adopting global shape descriptors corresponding to global shape entries which have a high number of votes.

10. The computer-based system of claim 5, further comprising:

means for generating a recognition signal from said selected retrieved entry, wherein said recognition signal is indicative of a recognized target object.

* * * * *