Diss. ETH No 9876

Discrete Voronoi Skeletons

A dissertation submitted to the

SWISS FEDERAL INSTITUTE OF TECHNOLOGY
ZURICH

for the degree of
Doctor of Technical Sciences

presented by

Robert L. Ogniewicz

Dipl. El. Eng. ETH
born the 3rd of October 1962
citizen of Frauenfeld (Thurgau)

Accepted on the recommendation of
Prof. Dr. O. Kübler, examiner
Prof. Dr. T. Pun, co-examiner

Zürich, 1992

Abstract

One of the strategic objectives of Computer Vision is the development of efficacious, flexible and *robust* methods for shape description. The main topic of this thesis is the presentation and discussion of a 2-dimensional shape representation, commonly known as *medial axis transform* (MAT) or *skeleton*. Intuitively, the medial axis transform can be interpreted as a reduction of a 2-D object to a graph-like structure, similar to a stick figure.

A concise definition of the skeleton in the continuum (grassfire analogy) was given by Blum (1967) [6]: Let us assume that the object is homogeneously covered with dry grass and then set its boundary on fire. The skeleton is defined as the locus where the fire fronts meet and quench.

For several reasons, the MAT has always been regarded as an attractive descriptor for planar shapes:

- The MAT is a region-based descriptor.
- According to its definition in the continuum, the MAT is invariant with respect to geometric transformations such as translation, rotation, and scaling.
- The MAT captures important visual cues such as symmetry, local width of the shape and shape complexity (branching points).
- The MAT can be used to describe "natural", "deformable" shapes.
- Modification of a boundary segment (e.g., due to occlusion) only influences a limited portion of the associated medial axis ("rich local support", according to Brady [11].
- The MAT incorporates adjacency and neighborhood information and thus can support grouping processes.

The most severe flaw of the MAT is its inherent sensitivity to details of the boundary. Moreover, attempts to transfer the original definition from the continuum to the raster grid often introduce additional drawbacks such as: Lack of Euclidean metrics, unduly fragmented skeletons, or extensive preprocessing as a prerequisite (e.g., robust polygonal approximation of the shape).

These drawbacks can be overcome by replacing the discrete distance map with the *Voronoi diagram* (VD) of boundary points (Chapter 2). The complete Voronoi diagram is equivalent to the medial axis of boundary points according to Blum's prairie fire concept. Since the computation (presented in detail in Appendix A) of the VD of a (discrete) set of boundary points is a central operation of our method, the resulting medial axis is henceforth denoted *discrete Voronoi medial axis* (DVMA).

Unfortunately, the DVMA is cluttered with fine-grain boundary information and has therefore to be *regularized.* Four so-called *residual functions* are proposed for the pruning of the DVMA (Chapter 3). These functions differ with respect to their individual computational complexity and their ability to remove spurious branches. Each edge of the DVMA is attributed a residual value, which expresses the expected stability of this edge with respect to noise, discretization effects, and geometric transformations.

The resulting *discrete Voronoi skeletons* (DVSK) are obtained by a simple threshold operation upon the attributed DVMA (Chapter 4). A reasonable threshold value is estimated by analyzing the effects of a variety of model artefacts. The DVSK is characterized by true Euclidean metrics and correct topology for a comfortably wide range of threshold values. The use of a vector-oriented rather than a grid-oriented encoding of the object's boundary as well as the introduction of rational numbers to describe the vertices of the VD makes the medial axis of thin structures appear much more congruous with its definition in the continuum than compared to its discrete counterpart. Moreover, it is possible

to achieve a perfectly symmetrical treatment of foreground and background. Thus both the skeletons of the foreground (endoskeletons) and the background (exoskeletons) can be computed simultaneously. Furthermore, the regularization concept is generalized for shapes described by multiple boundaries of arbitrary complexity. This allows to pass the symbolic rather than pixel-oriented output of an edge detector to the Voronoi skeleton apparatus.

The properties of the residual functions can be used for a hierarchical clustering of skeleton components, the *skeleton pyramid* (Chapter 6). Consequently, it is possible to extract a *main medial axis* which represents the most salient features of the underlying shape. Further analysis of the skeleton can then focus on skeleton fragments at lower levels of the hierarchy. Therefore, our regularization methods and the skeleton pyramid establish a *robust multiresolution skeletal representation of a shape.* The most important advantages of this structure are:

(1) It can capture the main characteristics of shapes with a significantly distorted, jagged boundary;

(2) The selection of the skeleton pertaining to a specific resolution level boils down to mere thresholding;

(3) The correspondence problem between different levels is avoided.

This clearly sets it apart from other approaches which first have to adapt the generating shape for each desired scale either by a blurring or smoothing operation and then require a recomputation of the skeleton for each resolution level.

Three distinct variants of accessing skeleton data are outlined in Chapter 5. First, an algorithm is introduced which implements the traversal of a discrete Voronoi skeleton. Second, we discuss how to "scan" a closed respectively nonclosed polygon of the skeleton. The third algorithm realizes a rotation about a particular skeleton vertex, thereby visiting all incident branches.

Performance figures presented in Appendix C illustrate that the evaluation and intelligent handling of complex data structures such as Voronoi skeletons or the skeleton pyramid is not necessarily a time consuming task. Several applications of the discrete Voronoi skeletons are outlined in Chapter 7. DVSK–based recognition of planar binary objects is illustrated by means of three distinct implemented algorithms and their results. While the first two methods aim at the recognition of geometrically rigid objects, the third algorithm is able to operate on scenes containing deformable shapes and a considerable amount of occlusion. Moreover, three additional applications are presented which embody or profit from our skeletonization method: Separation of abutting or overlapping binary objects prior to identification, extraction of line graphs, and document analysis.

Zusammenfassung

Ein strategisches Ziel der Bildverarbeitung ist die Entwicklung effizienter, flexibler und *robuster* Methoden zur Formbeschreibung und Formerkennung. Die vorliegende Arbeit stellt eine Methode zur Beschreibung flächiger Objekte vor, allgemein bekannt als *Mittelachsentransformation* (MAT) oder *Skelett.* Intuitiv läßt sich die MAT als die Reduktion eines zweidimensionalen Objekts auf einen Graphen interpretieren, vergleichbar mit einer Strichfigur. Eine einprägsame Definition des Skeletts im Kontinuum wurde 1967 von Blum [6] vorgeschlagen: Ein Objekt wird als homogene dürre Grasfläche dargestellt. Nachdem gleichzeitig entlang des Objektrands ein Feuer entfacht wurde, propagieren die Feuerfronten ins Objektinnere. Das Skelett wird als die Summe aller Punkte aufgefaßt, wo die Feuerfronten sich gegenseitig auslöschen. Diverse Eigenschaften der MAT lassen sie als interessanten Ansatz zur Formbeschreibung erscheinen:

- Die Mittelachsentransformation erfaßt die Flächencharakteristik eines Objekts
- Gemäß ihrer Definition im Kontinuum bleibt die MAT invariant gegenüber Operationen wie Translation, Rotation und Skalierung.
- Die MAT kodiert wichtige visuelle Charakteristika wie Symmetrie, lokaler Objektdurchmesser und Komplexität der Form (Verzweigungspunkte des Skeletts).
- Die MAT eignet sich zur Beschreibung "flexibler", "natürlicher" Formen.
- Die Modifikation eines Teils der Kontur verursacht eine Strukturveränderung der zugehörigen Mittelachse nur innerhalb einer begrenzten Zone ("rich local support", nach Brady [11]).
- Die MAT erfaßt topologische Nachbarschafts- und Adjazenzrelationen, sie kann deshalb zur Unterstützung von Gruppierungsprozessen herangezogen werden.

Eines der schwerwiegendsten Nachteile der Mittelachsenbeschreibung ist ihre Empfindlichkeit gegenüber Störungen. Ferner waren die meisten Versuche einer diskreten Implementierung der MAT von diversen Problemen begleitet: Verlust der euklidischen Metrik, fragmentierte Skelette oder aufwendige Vorverarbeitung (z.B. robuste Approximation eines Objekts durch Polygone).

Diese Nachteile lassen sich vermeiden, indem die diskrete Distanzkarte durch das *Voronoi–Diagramm* (VD) der Randpunkte ersetzt wird (Kapitel 2). Das vollständige Voronoi–Diagramm is äquivalent zur Mittelachsentransformation der Randpunktmenge im Sinne von Blum. Da die Berechnung des Voronoi–Diagramms (Details der verwendeten Implementierung sind im Anhang A dargestellt) eine zentrale Operation innerhalb unserer Skelettierungsmethode darstellt, bezeichnen wir die so entstehende MAT als *diskrete Voronoi–Mittelachse* (DVMA).

Leider enthält die DVMA unnötige Detailinformation, die zudem überaus empfindlich auf Rauschen und Diskretisierungseffekte reagiert. Die notwendige *Regularisierung* der DVMA wird durch die Einführung von vier sogenannten Residuen ermöglicht (Kapitel 3). Diese Residuen unterscheiden sich hinsichtlich ihrer Berechnungskomplexität und ihrer Fähigkeit, wenig signifikante Komponenten der Mittelachse zu unterdrücken. Jede Kante der DVMA wird mit dem entsprechenden Residuum gewichtet, das die erwartete Stabilität dieser Kante gegenüber Rauschen, Quantisierungseffekten und geometrischen Abbildungen ausdrückt.

Die in der DVMA enthaltenen diskreten Voronoi–Skelette (DVSK) können mittels einfacher Schwellwertbildung extrahiert werden (Kapitel 4). Ein sinnvoller Schwellwert läßt sich ermitteln, indem die Auswirkungen von Modell-Artefakten entlang der Kontur analysiert werden. Das DVSK ist durch echte euklidische Metrik und (für einen hin-

reichend großen Bereich der Schwellwerte) korrekte Topologie charakterisiert. Die vorliegende Implementation der Voronoi–Skelette basiert auf einer raster- anstatt vektororientierten Beschreibung der Objektberandung sowie rationaler Arithmetik zur Berechnung und Repräsentation des VD. Dadurch erhalten wir eine wesentlich präzisere Repräsentation der Mittelachse, insbesondere bei dünnen Strukturen. Ferner läßt sich eine völlig symmetrische Behandlung von Vordergrund und Hintergrund erreichen. Damit können zu gleicher Zeit Vordergrundskelett (Endoskelett) und Hintergrundskelett (Exoskelett) bestimmt werden. Das vorgestellte Regularisierungsverfahren läßt sich auf Objekte mit mehr als einer Kontur beliebiger Komplexität erweitern. Es ist deshalb möglich, das Resultat einer Kantendetektion in symbolischer Form als Ausgangsbasis für die Berechnung des DVSK zu verwenden.

Die Eigenschaften der Residuen lassen sich ferner zur hierarchischen Gruppierung der Skelettkomponenten verwenden (*Skelettpyramide*, Kapitel 6). Mithilfe dieser Datenstruktur ist es möglich, eine *Hauptmittelachse* zu definieren, die prägnante Merkmale einer Form erfaßt. Eine detailliertere Analyse des Objekts kann dann mittels Skelettkomponenten erreicht werden, die Teil einer niedrigeren Hierarchiestufe sind. Damit erlauben unser Regularisierungsverfahren und die Skelettpyramide eine *robuste skelettbasierte Formbeschreibung* im Sinne einer Auflösungspyramide. Die wichtigsten Vorteile dieser Struktur sind

(1) die Möglichkeit, ein sinnvolles Skelett für Objekte mit stark deformierter Berandung zu erhalten;

(2) die Extraktion des Skeletts zu einer bestimmten Auflösungsstufe geschieht durch bloße Schwellwertbildung; und

(3) es muß kein Korrespondenzproblem zwischen verschiedenen Auflösungsstufen gelöst werden.

Damit unterscheidet sich unsere Methode signifikant von anderen Ansätzen, bei denen zuerst die Objektberandung durch "Blurring" oder Glättung vorverarbeitet werden muß. Diese Verfahren verlangen die erneute Berechnung der gesamten Mittelachse für jede Auflösungsstufe.

Drei verschiedene Methoden, um auf die Skelettinformation zuzugreifen, werden in Kapitel 5 vorgestellt. Der erste Algorithmus dient zur Traversierung des gesamten Skelettgraphen. Ein zweiter Algorithmus erlaubt die Inspektion eines einzelnen (geschlossenen oder offenen) Voronoi–Polygons. Die dritte Methode implementiert eine Rotation um einen Skelettknoten, wobei alle die in den Knoten einlaufenden Kanten des DVSK besucht werden.

Die Leistungsfähigkeit der vorgestellten Skelettextraktion wird anhand einiger Beispiele im Anhang C illustriert. Die gemessenen Laufzeiten beweisen, daß unsere Methode durchaus mit existierenden rasterorientierten Verfahren konkurrieren kann. Einige Anwendungen der Voronoi–Skelette werden in Kapitel 7 skizziert. Wir beschreiben drei Varianten skelettbasierter Formerkennung für flächige binäre Objekte. Die ersten zwei Methoden zielen auf die Erkennung geometrisch starrer Formen ab. Das dritte Verfahren erlaubt die Identifikation "flexibler" Objekte in Szenen, die einen hohen Grad an Überlappungen aufweisen. Drei weitere Anwendungsgebiete, die sich der Mittelachsentransformation bedienen können, werden erwähnt: Separierung berührender oder überlappender Objekte ohne vorhergehende Erkennung, Beschreibung linienhafter Strukturen und Dokumentenanalyse.