

# "Active Shape Models" für die Verfolgung nicht-formfester Objekte in Farbbildfolgen

Andreas Koschan und Klaus Curio

Technische Universität Berlin

Institut für Technische Informatik

e-mail: koschan@cs.tu-berlin.de

<http://cv.cs.tu-berlin.de/mitarbeiter/koschan.html>

## Abstract

Das Auffinden und die Verfolgung (*tracking*) von Objekten in einer Bildsequenz gewinnen in letzter Zeit immer mehr an Bedeutung in der Bildanalyse. Besondere Anforderungen an die digitale Bildverarbeitung werden hierbei bei der Verfolgung von Objekten gestellt, deren Form (bzw. deren Silhouette) sich zwischen zwei Bildern verändert. Beispielsweise gehören Fußgänger in einer Straßenszene zu dieser Klasse von Objekten, die als nicht-formfeste Objekte (*non-rigid objects*) bezeichnet werden. In der Grauwertbildverarbeitung werden für die Verfolgung derartiger Objekte in einer Bildfolge z.B. sogenannte "Active Shape Models" (ASMs) eingesetzt. Eine Schwierigkeit bei ASMs bilden die Ausreißerprobleme speziell bei partiellen Objektverdeckungen. In dieser Arbeit wird eine Erweiterung der ASMs für Farbbilder angegeben und untersucht, inwieweit die Auswertung von Farbinformation zur Lösung dieser Problematik beitragen kann.

## 1 Einleitung

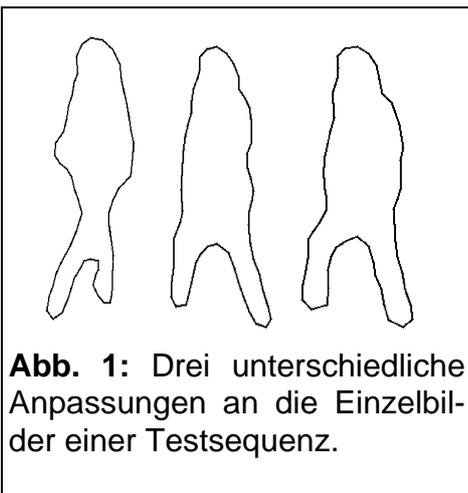
Für die Verfolgung von Objekten in Farbbildfolgen wurde von Brock-Gunn, Dowling und Ellis [1] die Verwendung von vierdimensionalen Schablonen vorgeschlagen. Wird jedoch die Beobachtung über einen langen Zeitraum und mit sehr vielen Einzelobjekten durchgeführt, so erhöht sich sowohl der Speicherplatzbedarf der Datenbank für die Schablonen als auch der Aufwand für die Suche einer Schablone in der Datenbank. „Active Shape Models“ [2] repräsentieren demgegenüber ein kompaktes Modell, dem durch eine Trainingsphase die Formvielfalt und die Farbverteilung einer Objektklasse beigebracht wird. Da sowohl die Form als auch die Farbverteilungen zwischen den in dieser Phase untersuchten Objekten korreliert vorliegen, wird je ein Satz unkorrelierter Merkmale berechnet, aus denen sich die Modellparameter ergeben. Hierdurch reduziert sich in der Regel die Anzahl der Modellparameter erheblich.

Ein sehr effizientes Verfahren zur Erkennung von farbigen Objekten stellt die Farbindexierung (*color indexing* [3]) dar. Anhand der Auswertung von Farbverteilungen erfolgt die Zuordnung zwischen dem Objekt im Bild und einem Objekt in der Datenbank. Dieses Verfahren benötigt in der Regel mehrere Ansichten des zu findenden Objektes (was beim Auffinden eines Passanten in einer Straßenszene nicht immer gewährleistet ist) und versagt teilweise bei partiellen Verdeckungen des Objektes. „Active Shape Models“ benötigen nicht mehrere Ansichten eines Objektes, da sie sich unter Ausnutzung von

Energiefunktionen an die jeweilige Silhouette eines im Bild dargestellten Objektes anpassen lassen. Eine Schwierigkeit bei diesen Modellen bilden jedoch die Ausreißerprobleme, die speziell bei partiellen Objektverdeckungen auftreten können. Nachfolgend wird eine Erweiterung der ASMs für Farbbilder angegeben und untersucht, inwieweit die Auswertung von Farbinformation zur Lösung der Ausreißerproblematik beitragen kann.

## 2 „Active Shape Models“

Die Idee der „Active Shape Models“ (ASMs) besteht darin, die Silhouette eines Objektes im Bild parameterisierbar zu modellieren, um anhand von zu bestimmenden Transformationen die sich in den Einzelbildern verändernden Silhouetten ineinander zu überführen. Abbildung 1 zeigt drei Silhouetten einer erfolgreichen Anpassung an das Objekt in den Einzelbildern. Ein ASM besteht aus einem automatisch erzeugten Punktverteilungsmodell (*point distribution model*), einer linearen Transformation und in dieser Arbeit zusätzlich aus der Einbeziehung der Farbinformation. Im Punktverteilungsmodell werden die Abstände charakteristischer Objektpunkte untereinander abgespeichert. Damit ist es möglich, gleichartige Objekte und deren Variationsmöglichkeiten in einem einzigen Modell zu beschreiben. Ein Punktverteilungsmodell repräsentiert die Formvielfältigkeit einer Objektklasse.



Das Punktverteilungsmodell wird aus der Durchschnittsform aller verwendeten Punktmengen und den Abweichungen der einzelnen Punkte jeder Punktmenge von dem korrespondierenden Punkt der Durchschnittsform gebildet. Die Punktmengen werden in einem  $2n$ -dimensionalen Merkmalvektor  $\mathbf{x}$  repräsentiert. Die Anzahl der Punkte wird mit  $n$  bezeichnet und die Anzahl der verwendeten Bilder mit  $N$ . Der Merkmalvektor  $\mathbf{x}_i$  des  $i$ -ten Bildes ist durch

$$\mathbf{x}_i = (x_{i0}, y_{i0}, x_{i1}, y_{i1}, \dots, x_{ik}, y_{ik}, \dots, x_{in-1}, y_{in-1})^T \text{ mit } i \in \{1, 2, \dots, N\}$$

gegeben. Für die automatische Ermittlung der Merkmalvektoren eines Objektes aus den Einzelbildern einer Videosequenz wird ein Verfahren nach Baumberg und Hogg [4] eingesetzt, welches eine Objekt-Hintergrund-Trennung benötigt. Aus Platzgründen wird an dieser Stelle auf die Darstellung der Hintergrundextraktion verzichtet.

## 3 Berechnung der Anpassungsparameter zweier Punktmengen

Zu gegebenen Punktmengen  $\mathbf{x}_1$  und  $\mathbf{x}_2$  wird eine Rotation  $\theta$ , eine Skalierung  $s$ , und eine Translation  $(t_x, t_y)$  gesucht, um  $\mathbf{x}_2$  durch diese geometrische Veränderung an  $\mathbf{x}_1$  bestmöglich anzupassen. Dazu wird die Energiefunktion (vgl. [2])

$$E = (\mathbf{x}_1 - \mathbf{M}(\mathbf{x}_2))^T \cdot \mathbf{W} \cdot (\mathbf{x}_1 - \mathbf{M}(\mathbf{x}_2)) \quad (1)$$

minimiert mit

$$\mathbf{M} \begin{pmatrix} x_{jk} \\ y_{jk} \end{pmatrix} = \begin{pmatrix} (s \cdot \cos\theta) \cdot x_{jk} - (s \cdot \sin\theta) \cdot y_{jk} + t_x \\ (s \cdot \sin\theta) \cdot x_{jk} + (s \cdot \cos\theta) \cdot y_{jk} + t_y \end{pmatrix}.$$

$\mathbf{W}$  ist eine Diagonalmatrix, deren Elemente eine Gewichtung für jeden Punkt darstellen. Wenn  $R_{kl}$  der Abstand zwischen den Punkten  $k$  und  $l$  eines Merkmalvektors und  $V_{R_{kl}}$  die Varianz dieser Abstände über alle Merkmalvektoren sind, kann die Gewichtung  $w_k$  für den  $k$ -ten Punkt durch

$$w_k = \left( \sum_{l=0}^{n-1} V_{R_{kl}} \right)^{-1}$$

bestimmt werden. Bei „stabilen“ Punkten sind die Varianzen relativ klein, wodurch deren Summe gering und die Gewichtung groß ist. Für die weitere Berechnung wird  $a = s \cdot \cos\theta$  und  $b = s \cdot \sin\theta$  angenommen. Um  $\theta$ ,  $s$  und  $\mathbf{t}$  so zu wählen, daß die Gleichung (1) minimiert wird, wird das folgende lineare Gleichungssystem gelöst:

$$\begin{pmatrix} X_2 & -Y_2 & W & 0 \\ Y_2 & X_2 & 0 & W \\ Z & 0 & X_2 & Y_2 \\ 0 & Z & -Y_2 & X_2 \end{pmatrix} \cdot \begin{pmatrix} a \\ b \\ t_x \\ t_y \end{pmatrix} = \begin{pmatrix} X_1 \\ Y_1 \\ C_1 \\ C_2 \end{pmatrix}, \text{ mit}$$

$$X_i = \sum_{k=0}^{n-1} w_k x_{ik}, \quad Y_i = \sum_{k=0}^{n-1} w_k y_{ik}, \quad W = \sum_{k=0}^{n-1} w_k, \quad Z = \sum_{k=0}^{n-1} w_k (x_{2k}^2 + y_{2k}^2),$$

$$C_1 = \sum_{k=0}^{n-1} w_k (x_{1k} x_{2k} + y_{1k} y_{2k}) \quad \text{und} \quad C_2 = \sum_{k=0}^{n-1} w_k (x_{2k} y_{1k} - x_{1k} y_{2k}).$$

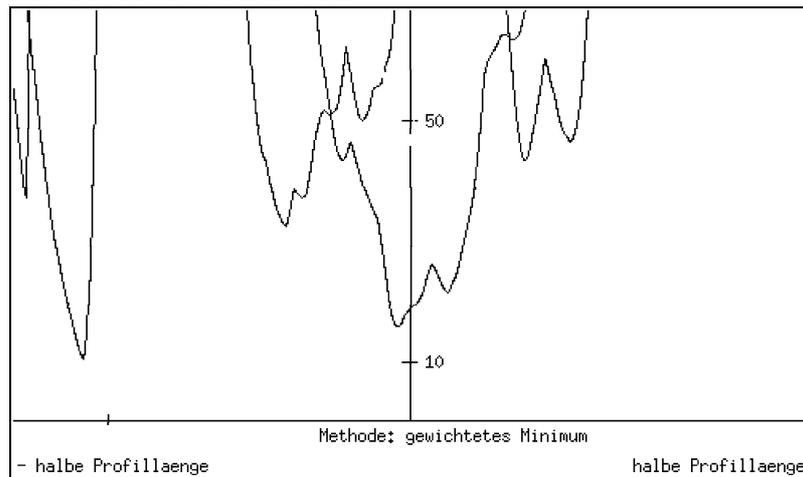
Die Lösungen für  $a$ ,  $b$ ,  $t_x$  und  $t_y$  können zum Beispiel mit Hilfe des Gauss-Jordan-Verfahrens bestimmt werden.

## 4 Erweiterung des Verfahrens für Farbbildsequenzen

Bei der Grauwertvariante werden die Energiefunktionen entlang der Normalen für einen repräsentativen Punkt in der Grauwertverteilung bestimmt. Diese Vorgehensweise läßt sich auf Farbbilder erweitern, indem zunächst für jede Komponente der Farbvektoren separat die Energiefunktionen bestimmt werden. Bei der Minimumsanalyse müssen dann die so erhaltenen Minima zu einem „gemeinsamen“ Minimum verknüpft werden. Eine Möglichkeit besteht darin, das jeweils kleinste Minimum in den drei Farbkomponenten als Kandidat auszuwählen. Falls jedoch in einem der drei Farbkanäle ein „Ausreißer“ auftritt (vergleiche Abb. 2), so würde unter Umständen dieser Ausreißer ausgewählt.

Eine andere Vorgehensweise besteht darin, den Mittelwert der absoluten Minima in allen drei Farbkomponenten auszuwählen. Ein Ausreißer in einem Farbkanal führt jedoch auch in diesem Fall zu einem falschen Ergebnis. Weiterhin stellt unter Umständen der Mittelwert einen Wert dar, der mit keiner der betrachteten Energiefunktionen korrespondiert. Als beste Variante hat sich bei den von uns untersuchten Bildsequenzen die Bestimmung des Medians der absoluten Minima in den drei Farbkanälen herausgestellt. Hierdurch wird der Einfluß eines Ausreißers in den Minima der

Energiefunktionen minimiert. Bei der Anpassung der Silhouetten aneinander können jedoch weiterhin falsche Werte auftreten.



**Abb. 2:** Die Energiefunktionen für die drei Farbkomponenten mit einem Ausreißer in der roten Komponente.

## 5 Ergebnisse

Die Testbildsequenzen wurden mit einer Videokamera mit der Hand aufgenommen und weisen gewisse Bewegungsunschärfen auf. Es wurden 11 Farbbildsequenzen mit jeweils 142 Einzelbildern untersucht. Die Auswahl des in der Bildsequenz zu verfolgenden Objektes erfolgt anhand eines Verfahrens nach Baumberg und Hogg [4]. In den Testsequenzen bewegt sich jeweils nur ein Objekt. Falls sich mehrere Objekte in der Szene bewegen, so muß für jedes Objekt ein ASM berechnet werden. Sind die Formen der zu verfolgenden Objekte hinreichend unterschiedlich (z.B. ein Hund, ein Ball und eine Person), so ist eine Verfolgung der Objekte unter Verwendung von ASMs möglich.

Bewegen sich jedoch mehrere Personen in der Bildsequenz, so würde eine Verfolgung ausschließlich anhand der Silhouette im Bild Schwierigkeiten bereiten. Ein Lösungsansatz hierfür bestünde in einer gewichteten Kombination eines die Form repräsentierenden Verfahrens (z.B. mit ASMs) mit einem die Farbverteilung repräsentierenden Verfahren (z.B. Farbindizierung). Hierdurch könnte unterschieden werden, ob es sich um mehrere Objekte gleicher Form aber unterschiedlicher Farbverteilung oder um mehrere Objekte gleicher Farbverteilung aber unterschiedlicher Form handelt.

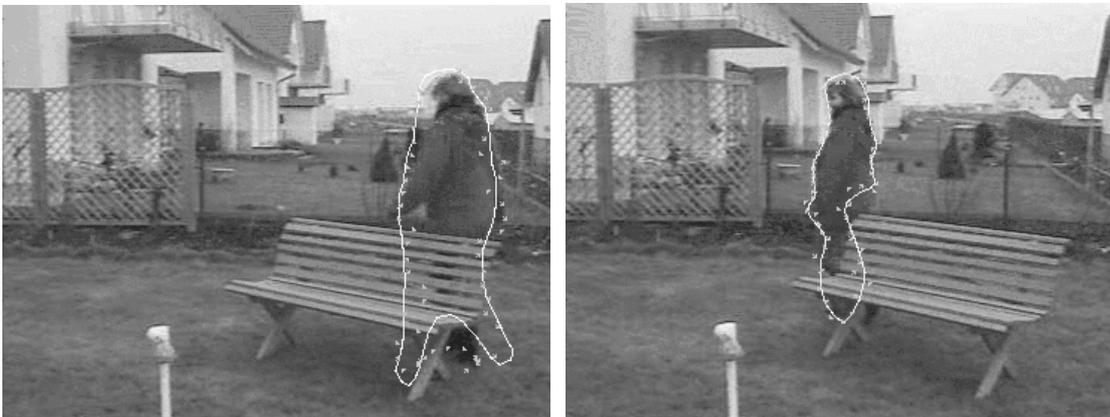
Zur Initialisierung der Durchschnittsform des zu verfolgenden Objektes gehört auch die Festlegung der Formparameter für das Punktverteilungsmodell. Diese wurden in der vorliegenden Arbeit manuell eingegeben. Von Hill, Taylor und Cootes [5] wurde hierfür ein genetischer Algorithmus vorgeschlagen, der aus einer zufällig festgelegten Menge von Startwerten die „besten“ Formparameter bestimmt. Dieses rechenzeitaufwendige Verfahren wurde von uns bisher nicht untersucht.

Zur Begrenzung der Rechenzeit wird weiterhin eine maximal zulässige Verschiebung des Objektes zwischen zwei Bildern festgelegt. Dieser Wert wurde bei uns auf 15 Pixel begrenzt. Abb. 3 zeigt links das erste Bild einer Testsequenz und rechts die Initialisierung der Durchschnittsform (vgl. Farbabb. 1). Das Ergebnis der Anpassung im ersten Bild und der Modellzustand für Bild 19 ist in Abb. 4 dargestellt (vgl. Farbabb. 2). Abb. 5 zeigt die

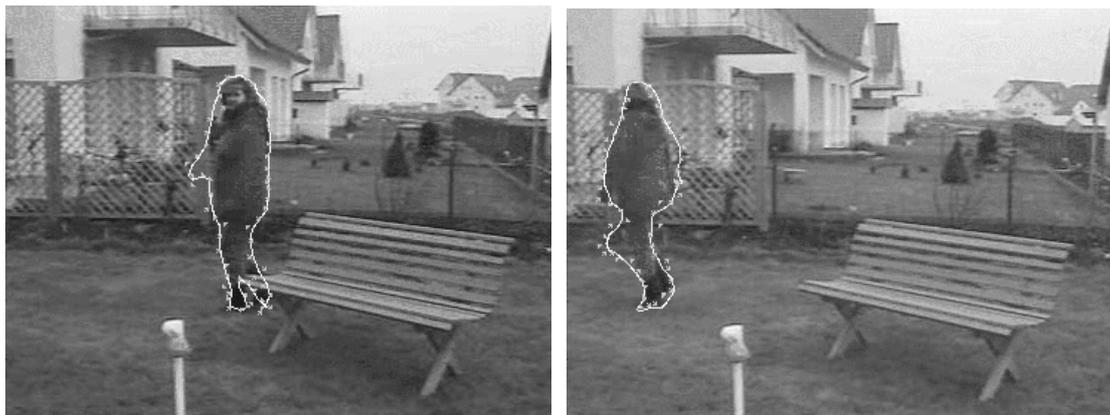
Modellzustände in den Bildern 27 und 33 (vgl. Farbabb. 3). Ein Beispiel für 18 vom Hintergrund isolierte Objekte wird in Farbabb. 4 dargestellt



**Abb. 3:** Links: Erstes Bild einer Testsequenz. Rechts: Initialisierung der Durchschnittsform (vgl. Farbabb. 1).



**Abb. 4:** Links: Ergebnis der Anpassung im ersten Bild. Rechts: Modellzustand für Bild 19 (vgl. Farbabb. 2).



**Abb. 5:** Links: Modellzustand für Bild 27. Rechts: Modellzustand für Bild 33 (vgl. Farbabb. 3).

## 6 Zusammenfassung

Es wurde ein Verfahren vorgestellt, mit dem eine laufende Person in einer Videosequenz verfolgt werden kann. Hierfür wurde das „Active Shape Model“ auf Farbbilder erweitert. Für die Kombination der Information aus den einzelnen Komponenten der

Farbbildvektoren erwies sich die Medianberechnung der Minima in den Energiefunktionen als günstig. Das Verfolgungsverfahren benötigt (außer zur Initiierung des zu verfolgenden Objektes) keine statische Kamera. Es zeigte sich auch gegenüber partiellen Objektverdeckungen relativ robust. Für eine statistische Aussage ist die Einbeziehung weiterer Testbildsequenzen notwendig. Das Ausreißerproblem konnte durch die Auswertung von Farbinformation nur teilweise gelöst werden, was vor allem durch die Korreliertheit der Farbkanäle begründet ist. Ziel zukünftiger Forschungsarbeiten ist eine verfeinerte Analyse der Energiefunktion, bei der auch die Nachbarn eines Punktes berücksichtigt werden. Hierdurch könnte die Anzahl der Ausreißer weiter reduziert werden. Weiterhin ist eine Kombination eines die Form repräsentierenden Verfahrens (z.B. mit ASMs) mit einem die Farbverteilung repräsentierenden Verfahren (z.B. Farb-indizierung) geplant.

## 7 Danksagung

Unser Dank gilt der Technischen Universität Berlin an der diese Arbeiten durchgeführt wurden.

## 8 Literatur

- [1] S.A. Brock-Gunn, G.R. Dowling, and T.J. Ellis. Tracking using colour Information. In *Third International Conference on Automation, Robotics and Computer Vision ICARCV '94*, Singapore, pages 686-690, 1994.
- [2] T.F. Cootes, D.H. Cooper, C.J. Taylor, and J. Graham. Active Shape Models - Their Training and Application. In *Computer Vision and Image Understanding 61*, pages 38-59, 1995.
- [3] M.J. Swain and D.H. Ballard. Color indexing. In *Int. Journal of Computer Vision 7*, pages 11-32, 1991.
- [4] A. Baumberg and D. Hogg. Learning flexible models from image sequences. In *3rd European Conference on Computer Vision ECCV'94*, Stockholm, pages 299-308, 1994.
- [5] A. Hill, C.J. Taylor, and T.F. Cootes. A Generic System for Image Interpretation Using Flexible Templates. In *European Conference on Computer Vision ECCV'94*, Stockholm, pages 276-285, 1994.



**Farbabbildung 1:** Links: Erstes Bild einer Testsequenz. Rechts: Initialisierung der Durchschnittsform.



**Farbabbildung 2:** Links: Ergebnis der Anpassung im ersten Bild. Rechts: Modellzustand für Bild 19.



**Farbabbildung 3:** Links: Modellzustand für Bild 27. Rechts: Modellzustand für Bild 33.



**Farbabbildung 4:** 18 vom Hintergrund isolierte Objekte.