

Bewegungsanalyse in Bildfolgen auf Basis eines n:m-Matchings von Farbregionen

B. Melzer

A. Miene

Th. Hermes

TZI Technologie-Zentrum Informatik, Fachbereich Mathematik und Informatik,
Universität Bremen

Postfach 330440, D-28334 Bremen / Universitätsallee 21-23 D-28359 Bremen

e-mail: bmelzer@tzi.org, andrea@tzi.org, hermes@tzi.org

Zusammenfassung

Im Rahmen dieser Arbeit wird ein merkmalsbasiertes Bewegungsanalyseverfahren für Bildfolgen auf der Grundlage von korrespondierenden Farbregionen vorgestellt. Es lässt sich aufteilen in fünf Verarbeitungsschritte: Vorverarbeitung, Farbsegmentierung, Merkmalsberechnung, Matching und Bewegungsschätzung. Die ersten drei Verarbeitungsschritte werden dabei auf die Einzelbilder angewendet, während das Matching Korrespondenzen zwischen jeweils zwei Bildern sucht. Die Bewegungsschätzung schließlich ermöglicht, Farbregionen über eine größere Anzahl von Bildern zu verfolgen.

1 Einleitung

Die Untersuchung der Veränderung von zeitlich benachbarten oder nahe beieinander liegenden Bildern ermöglicht die Bestimmung der Bewegung von Bildbereichen und lässt über die Analyse der Einzelbilder hinaus weitere Rückschlüsse auf den Inhalt der Bildfolge zu. Anwendungsbereiche von Bewegungsanalyseverfahren sind die Verfolgung von Objekten, Robotersteuerung, Fahrzeugnavigation oder die bewegungskompensierte Kodierung von Videos zur Kompression für die Speicherung oder Übertragung von Multimediaten.

Das hier vorgestellte Verfahren gehört zu den merkmalsbasierten Bewegungsanalyseverfahren, bei denen die Zuordnung nicht pixelbasiert, sondern anhand von Merkmalen - in diesem Fall Merkmalen segmentierter Farbregionen - erfolgt. Das Verfahren besteht aus fünf Verarbeitungsschritten: Vorverarbeitung, Farbsegmentierung (Abschnitt 2), Merkmalsberechnung (Abschnitt 3), Matching (Abschnitt 4) und Bewegungsschätzung (Abschnitt 5).

Die Bewegung von Farbregionen wird durch *Verschiebungsvektoren* beschrieben. Damit wird eine Grundlage geliefert, die Rückschlüsse auf Bewegungen von tatsächlichen Objekten in natürlichen Szenen ermöglichen kann. Diese Bestimmung von dreidimensionaler Bewegung wird in dieser Arbeit jedoch nicht näher betrachtet.

Einige Anmerkungen zu Experimenten mit der Implementierung des entwickelten Verfahrens sowie eine Zusammenfassung enthält Abschnitt 6.

2 Farbsegmentierung

Das Ziel jeder Segmentierung ist, das Bild in zusammenhängende Segmente (auch Regionen genannt) zu unterteilen, die jeweils in bezug auf die zugrundeliegende Eigenschaft – in diesem Fall der Farbe – homogen, untereinander aber visuell unterschiedlich sind.

Da die Farbregionen die Grundlage für die weiteren Verarbeitungsschritte bilden, ist eine gute Qualität der Segmentierungsergebnisse eine wichtige Voraussetzung für gute Ergebnisse der Bewegungsanalyse. Bei der Aufnahme von digitalen Bildern kommt es häufig zu Rauschen, was durch relativ starke Veränderungen der Farb- und Helligkeitswerte von nahe beieinander liegenden Bildpunkten auch in homogenen Bereichen bemerkbar macht. Das Rauschen wird im Rahmen der Vorverarbeitung mittels Standardglättungsverfahren wie Mittelwertfilter und Median-Filter [Abm94] unterdrückt, wodurch robustere Ergebnisse der anschließenden Farbsegmentierung ermöglicht werden.

Insbesondere in stark texturierten Bereichen eines Bildes, die in bezug auf die Farbeigenschaften sehr inhomogen sind, sind die Grenzen einer Farbsegmentierung erreicht. Aus diesem Grund erlaubt das entwickelte Farbsegmentierungsverfahren, daß Bildbereiche, die den Homogenitätskriterien nicht entsprechen, unsegmentiert bleiben dürfen.

Neben der Qualität der Ergebnisse und Robustheit der Verfahren ist gerade im Bereich der Videoanalyse auch das Laufzeitverhalten der Farbsegmentierung von großer Bedeutung. Die meisten bekannten Farbsegmentierungsverfahren (einen Überblick bietet [SK94]) wurden für die Segmentierung von Einzelbildern entwickelt. Besonders rechenintensive Verfahren sind für die Videoanalyse nicht geeignet.

Das entwickelte Farbsegmentierungsverfahren basiert auf einem schnellen zentroiden Bereichswachstum. Bereichswachstumsverfahren gehen von kleinen Bildbereichen, sogenannten initialen Zellen, aus und verketteten benachbarte Bildpunkte mit der Region, die einem Ähnlichkeitskriteriums entsprechen. Bei einer zentroiden Verkettung wird hierzu der Mittelwert der Region zum Vergleich herangezogen. Sie liefert dadurch robustere Ergebnisse als z.B. die einfache Verkettung, die benachbarter Bildpunkte vergleicht, wodurch es bei langsamen Farbübergängen zu falschen Verschmelzungen von Regionen kommen kann. Bei der zentroiden Verkettung beeinflusst die Reihenfolge, in der die Bildpunkte bearbeitet werden, das Ergebnis. Je kleiner die Region noch ist, umso wichtiger ist die Reihenfolge; besonders groß ist daher der Einfluss des Startpunktes. Zwei unterschiedliche Startpunkte liefern in der Regel unterschiedliche Segmentierungsergebnisse ([Zuc76]). Besonders geeignet sind Startpunkte, die in homogenen Bildbereichen liegen, da dann bereits eine relativ große Menge Bildpunkte zur Berechnung eines stabilen Mittelwertes beigetragen hat, bevor die (inhomogeneren) Randbereiche der Region erreicht werden. Einzelne Bildpunkte, die schon stärker abweichen, jedoch der Region noch zugeordnet werden, können den Mittelwert nun nicht mehr so stark beeinflussen. Aus diesem Grund werden für das Bereichswachstumsverfahren Startpunkte bestimmt, die in einer homogenen Umgebung bestimmter Größe und Qualität liegen.

Ein besonderer Vorteil der Farbsegmentierung besteht außerdem darin, dass die Segmentierung je nach Anforderungen an die Ergebnisse wahlweise in einem der Farbräume ([Ric81]) RGB, HLS, CIE-Lab und CIE-Luv ([CIE71]) durchgeführt werden kann.

Das Verfahren arbeitet wie folgt: Zunächst wird das Bild ggf. in den zu verwendenden Farbraum umgerechnet. Das Bild wird zeilenweise von links nach rechts und von oben nach unten abgearbeitet. Für jeden betrachteten Bildpunkt (m, n) wird, sofern er noch keiner Region zugeordnet worden ist, überprüft, ob er sich in einer homogenen quadratischen Umgebung der Größe $(\Delta r \times \Delta r)$ befindet. Die Umgebung des Bildpunktes gilt als homogen, wenn die Standardabweichung zwischen dem Punkt und den Punkten in der Umgebung einen gegebenen Schwellwert nicht überschreitet. Die Größe der Umgebung sowie der Schwellwert sind variabel. Ist der Punkt in einer homogenen Umgebung, wird er als Startpunkt für das Bereichswachstumsverfahren markiert. Dann wird von den Start-

punkten ausgehend parallel mittels eines rekursiven Verfahrens ein Bereichswachstum mit zentroider Verkettung wahlweise in vier oder acht Richtungen vorgenommen. Dabei wachsen die Regionen, solange benachbarte, noch nicht segmentierte Bildpunkte einen Farbabstand zu ihrem aktuellen mittleren Farbwert haben, der unterhalb eines variierbaren Schwellwerts liegt. Im Detail ist das Verfahren in [Mel02] beschrieben.

Es hat sich als sinnvoll herausgestellt, kleine Regionen in einem an die Segmentierung anschließenden Verarbeitungsschritt mit benachbarten Regionen zusammenzufassen. Dazu kann eine Mindestgröße für Regionen vorgegeben werden, so dass Regionen unterhalb der Mindestgröße mit benachbarten Regionen verschmelzen. Daraus ergibt sich insbesondere auch eine maximale Anzahl Regionen in einem Bild, die den Rechenaufwand für die weiteren Verarbeitungsschritte besser abschätzbar macht.

Abbildung 1 zeigt den Vorteil einer Auswahl von Startpunkten in homogener Umgebung gegenüber der Verwendung beliebiger Startpunkte.

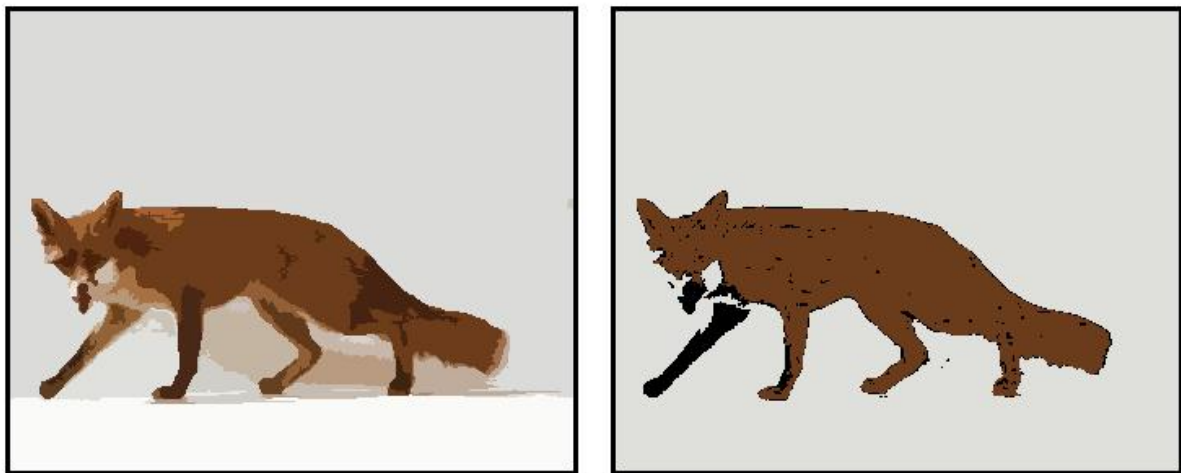


Abbildung 1: Beispiele für die Farbsegmentierung: links im RGB Farbmodell ohne Homogenitätsüberprüfung, rechts im CIE-Lab Farbraum mit Homogenitätsüberprüfung der Startpunkte (11×11 Umgebung). Hier hat die Farbsegmentierung lediglich zwei Regionen geliefert – die schwarz eingefärbten Bereiche blieben unsegmentiert.

3 Merkmalsberechnung

Die *Merkmalsberechnung* bestimmt für jede der Farbregionen eine Menge von Merkmalen – jede Farbregion wird im Weiteren durch diese Merkmale beschrieben und nicht mehr durch eine Menge von Bildpunkten.

Das auf der Basis der berechneten Merkmale arbeitende Matchingverfahren findet nicht nur Korrespondenzen zwischen einzelnen Regionen, sondern ist in der Lage, Übereinstimmungen zwischen Regionengruppen aus zwei Bildern zu finden. Aus diesem Grund werden Merkmale gewählt, die sich für eine Regionengruppe leicht aus den Merkmalen der Einzelregionen bestimmen lassen. Als geeignete Merkmale haben sich die in [RR97] vorgeschlagenen Merkmale erwiesen: mittlerer Farbwert, Größe, Schwerpunkt, kleinstes umschließendes Rechteck, sowie Proportion und Orientierung jeder Farbregion.

Der mittlere Farbwert wird bereits während der Farbsegmentierung berechnet. Die übrigen Merkmale beschreiben geometrische Eigenschaften der Regionen, die sich mit Ausnahme des kleinsten umschließenden Rechtecks alle aus den Momenten m_{pq} bis zur zweiten Ordnung ([Abm94]) bestimmen lassen. Die Momente wiederum lassen sich für eine Gruppe von Regionen leicht durch Addition der Momente der Einzelregionen bestimmen. Die Größe einer Region wird durch die Anzahl der Bildpunkte (also das Moment m_{00}), die Schwerpunktkoordinaten durch Division der beiden Momente erster Ordnung (m_{01} bzw. m_{10}) jeweils durch m_{00} bestimmt ([Abm94]). Als Proportion wird das Verhältnis zwischen der Länge der Haupt- und der Nebenachse der Ellipse mit den äquivalenten normalisierten Zentralmomenten zweiter Ordnung verwendet. Als Orientierung einer Region wird die Orientierung der Hauptachse dieser Ellipse verwendet. Die Berechnung der Längen der Haupt- und Nebenachse sowie der Orientierung ist in [HS92] beschrieben. Das kleinste umschließende Rechteck wird beschrieben durch die obere linke und die untere rechte Ecke, indem jeweils die kleinste und größte x- und y-Koordinaten aller zu einer Region gehörenden Bildpunkte ermittelt werden.

4 Matching

Das Ziel des *Matching* ist eine Zuordnung von Mengen von Farbregionen in Bildpaaren anhand ihrer Merkmale. Dabei werden nicht nur Korrespondenzen zwischen einzelnen Farbregionen zweier Einzelbilder gesucht, sondern basierend auf einem Ansatz von Rehrmann und Rothhaar ([RR96]) auch Übereinstimmungen zwischen n Regionen in einem Bild und m Regionen im zweiten Bild ($n:m$ -Matching).

Zunächst ist zu jedem einzelnen Merkmal ein Ähnlichkeitsmaß zum Vergleich der Regionen zu definieren. Die Ergebnisse werden auf einen Bereich zwischen 0 und 1 normalisiert. Dabei bedeutet ein Ähnlichkeitsmaß von 1, dass die Merkmale gleich sind und 0, dass der Unterschied zwischen den Merkmalen den maximal möglichen Wert annimmt.

- Farbe: $EQ_{color} = 1 - \frac{\Delta E}{\Delta E_{max}}$ mit ΔE : Farbabstand im verwendeten Farbraum, ΔE_{max} maximal möglicher Farbabstand im verwendeten Farbraum.
- Größe: $EQ_{area} = 1 - \frac{|m_{00_0} - m_{00_1}|}{m_{00_0} + m_{00_1}}$
- Schwerpunkt: $EQ_{centre} = 1 - \frac{\sqrt{(x_{s_0} - x_{s_1})^2 + (y_{s_0} - y_{s_1})^2}}{\sqrt{M^2 \cdot N^2}}$ mit x_{s_i}, y_{s_i} Schwerpunktkoordinaten; M, N Breite, Höhe des Bildes.
- Proportion: $EQ_{prop} = 1 - \frac{\left| \frac{a_0}{b_0} - \frac{a_1}{b_1} \right|}{\frac{a_0}{b_0} + \frac{a_1}{b_1}}$ mit a_i, b_i Längen der Haupt- und Nebenachsen der Ellipse.
- Orientierung: $EQ_{ori} = 1 - \left(\frac{\Delta \theta}{90} \left(1 - \frac{2}{\frac{a_0}{b_0} + \frac{a_1}{b_1}} \right) \right)$ mit $\Delta \theta$ Differenz der Orientierung der beiden Ellipsenhauptachsen (im Gradmaß).
- Kleinstes umschließendes Rechteck: $EQ_{mbr} = 1 - \frac{1}{2} \left(\frac{|w_0 - w_1|}{w_0 + w_1} + \frac{|h_0 - h_1|}{h_0 + h_1} \right)$ mit w_i, h_i Breite und Höhe des kleinsten umschließenden Rechtecks.

Die Orientierung ist abhängig von der Proportion ein unterschiedlich aussagekräftiges Merkmal. Für einen Kreis ist die Orientierung vollkommen irrelevant, ihre Bedeutung nimmt mit wachsender Proportion zu. Deshalb wird bei der Berechnung der Orientierungsähnlichkeit mit einem Gewichtungsfaktor multipliziert, der von der Proportion der betrachteten Regionen abhängt und sicherstellt, dass der resultierende Orientierungsunterschied sich 0 annähert, je bedeutungsloser die Orientierung ist.

Nachdem alle verwendeten Merkmale und deren Ähnlichkeit für die Regionen bzw. Regionengruppen berechnet worden sind, kann nun die Gesamtähnlichkeit bestimmt werden. Sie ergibt sich aus der gewichteten Summe von Ähnlichkeiten der einzelnen Merkmale. Für jedes Merkmal kann dazu ein Gewichtungsfaktor ω festgelegt werden, wobei die Summe aller Gewichtungsfaktoren ω genau 1 ergibt.

4.1 $n : m$ -Matching

Beim $n : m$ -Matching werden Teilmengen von Regionen aus den beiden segmentierten Bildern bestimmt, die zueinander passen. Praktisch alle Farbsegmentierungsverfahren liefern für natürliche Videoszenen teilweise unter- oder übersegmentierte Ergebnisbilder. Während Untersegmentierung insbesondere bei sich unterschiedlich bewegendem Regionen einen unwiederbringlichen Informationsverlust darstellt, kann eine Übersegmentierung mit dem $n : m$ -Matching ausgeglichen werden. Die Merkmale sind so ausgewählt worden, dass es einfach möglich ist, die Merkmale für die Vereinigung mehrerer Regionen zu berechnen.

Es kann jedoch nicht für sämtliche möglichen Regionenteilmengen aus zwei Bildern ein Matching-Versuch unternommen werden, weil der Aufwand hierfür exponentiell mit der Anzahl der Regionen wächst. Um ein $n : m$ -Matching mit praxistauglichem Laufzeitverhalten zu erzielen, ist eine weitere Reduzierung der zu betrachtenden Teilmengen notwendig. Einige Möglichkeiten hierzu werden in [RR96] vorgestellt. Das im Rahmen des hier vorgestellten Verfahrens eingesetzte Matching fasst nur solche Regionen zu Teilmengen zusammen, die ähnliche Farbwerte und eine bestimmte räumliche Nähe haben (vgl. [Mel02]). Ferner wird die maximale Anzahl von Regionen einer Teilmenge eingeschränkt. Abbildung 2 zeigt ein Beispiel für einen $2 : 2$ -Match. Durch teilweise Verdeckung wird ein sonst homogener Bereich in zwei untersuchten Bildern in jeweils zwei Farbregionen aufgeteilt. Beim $n : m$ -Matching liefert der dargestellte $2 : 2$ -Match eine höhere Gesamtähnlichkeit der Merkmale als jede Kombination der beteiligten Einzelregionen.

5 Bewegungsschätzung

Bisher wurde lediglich die Verschiebung von Farbregionen zwischen Bildpaaren untersucht. Für jede gefundene Korrespondenz zwischen den Regionen zweier Bilder wird dabei aus der Veränderung der Position die Bewegungsrichtung und die Bewegungsgeschwindigkeit bestimmt. Der so berechnete *Bewegungsvektor* und die Merkmale der an der Korrespondenz beteiligten Regionen werden genutzt, um die Ergebnisse des Matching zusammenzufassen und so eine *Bewegungsschätzung* der sich bewegendem Regionen über eine längere Sequenz von Einzelbildern hinweg vorzunehmen. Dazu werden die beim Matching zwischen den Bildpaaren ermittelten korrespondierenden Regionengruppen einer Liste sich bewegendem Bildbereiche zugeordnet, die bereits aus früheren Einzelbildern bekannt sind. Dabei beschreiben die einzelnen Einträge dieser Liste die Be-

wegungsschritte einer Region oder Regionengruppe über mehrere Einzelbilder. Wird aus den Ergebnissen des Matching auf der Grundlage der Merkmale und der Bewegungsvektoren eine passende Korrespondenz für eine bereits bekannte Bewegung gefunden, wird sie als ein neuer Bewegungsschritt hinzugefügt. Korrespondenzen die keiner bekannten Bewegung zugeordnet werden konnten, bilden den ersten Bewegungsschritt einer neuen Bewegung.

6 Experimente

Das Verfahren wurde anhand verschiedener synthetischer und natürlicher Bildfolgen getestet. Der Vorteil synthetisch erzeugter Videos liegt darin, dass alle Merkmale der im Bild vorkommenden Farbreionen bereits zu Anfang bekannt sind und die Ergebnisse des getesteten Verfahrens somit objektiv überprüfbar sind. Die Tests mit synthetisch erzeugten Videos haben gezeigt, dass die Merkmale der Regionen sehr genau mit den bei der Erzeugung des Videos verwendeten Parametern übereinstimmen. Die Bewegung der Regionen wird in diesen Bildfolgen richtig erkannt und die Regionen werden zuverlässig verfolgt.

Auch in den natürlichen Videosequenzen konnten zu einzelnen Objekten gehörende Gruppen von Farbreionen zuverlässig verfolgt werden. Eine besonders robuste Bewegungserkennung ist möglich, wenn die zu verfolgenden Bereiche sich farblich deutlich von ihrer Umgebung abheben. Verschmelzen Objektregionen aufgrund von zu geringen Farbunterschieden mit ihrem Hintergrund, dann gehen zwar Regionen „verloren“. Diese können aber einige Bilder später wiedergefunden werden, falls die Regionen des verfolgten Objektes wieder richtig segmentiert werden.

Außerdem ergaben die Experimente, dass Bewegung umso zuverlässiger erkannt werden kann, je größer die zu verfolgenden Farbreionen sind. Für sehr kleine Regionen ist das Verfahren weniger gut geeignet, da in diesem Fall die Merkmale zwischen aufeinander folgenden Bildern stark schwanken und Korrespondenzen nicht zuverlässig gefunden werden können.

Literatur

- [Abm94] ABMAYR, WOLFGANG: *Einführung in die digitale Bildverarbeitung*. Teubner Verlag, Stuttgart, 1994.

- [CIE71] CIE (COMMISSION INTERNATIONALE DE L'ÉCLAIRAGE): *Colorimetry. Colorimétrie. Farbmessung. Publ. CIE Nr. 15*, 1971.

- [HS92] HARALICK, ROBERT M. and LINDA G. SHAPIRO: *Computer and Robot Vision*. Addison-Wesley Publishing Company, Reading, MA, USA, 1992.

- [Mel02] MELZER, BJÖRN: *Bewegungsanalyse in Bildfolgen auf Basis von Farbreionen*. Diplomarbeit, Universität Bremen, 2002.

- [Ric81] RICHTER, MANFRED: *Einführung in die Farbmetrik*. de Gruyter, Berlin, New York, Zweite Auflage, 1981.

- [RR96] REHRMANN, VOLKER und M. ROTHHAAR: *Objektorientierte Bewegungsschätzung in Farbbildfolgen*. In: *2. Workshop Farbbildverarbeitung*, Seiten 69–72, 1996. Ilmenau, 10.-11. Oktober.
- [RR97] REHRMANN, VOLKER and MARTIN ROTHHAAR: *Detection and Tracking of Moving Objects in Color Outdoor Scenes*. In *Proceedings: 30th ISATA (International Symposium on Automotive Technology and Automation), Dedicated Conference on Robotics, Motion and Machine Vision in the Automotive Industry, Florence, 16-19th June, 1997*.
- [SK94] SKARBEK, WLADYSLAW and ANDREAS KOSCHAN: *Colour Image Segmentation – A Survey*. Technical Report 94-32, TU Berlin, 1994.
- [Zuc76] ZUCKER, STEVEN W.: *Region growing: Childhood and adolescence*. *Computer Graphics and Image Processing*, 5:382–399, September 1976.

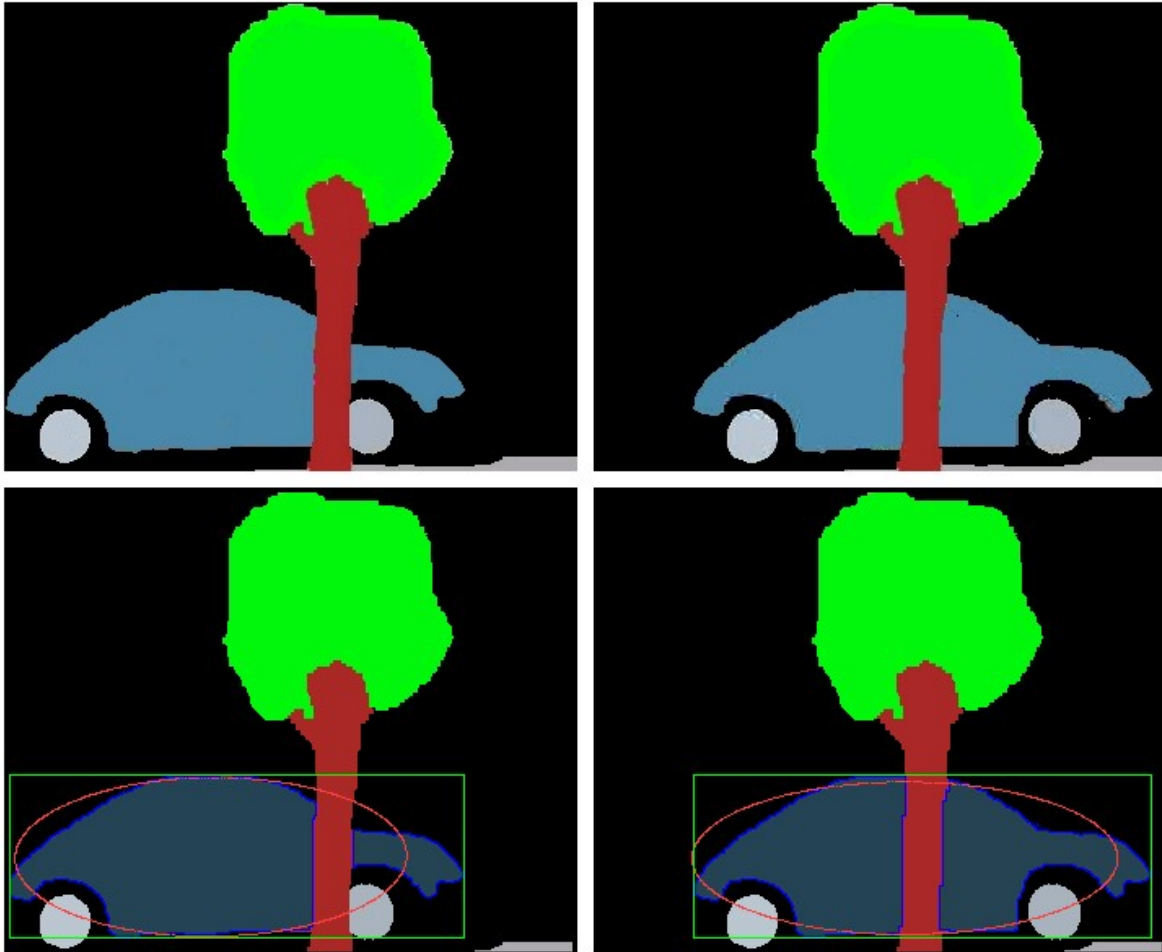


Abbildung 2: Einfaches Beispiel für das $n : m$ -Matching: Zwei Einzelbilder (Originale oben) aus einem synthetisch erzeugten Video mit einem gefundenen $2 : 2$ -Match der blau umrandeten Regionen (unten). Die Ellipse mit den äquivalenten normalisierten Zentralmomenten zweiter Ordnung (rot), aus deren Parametern die Merkmale Proportion und Orientierung bestimmt werden und das kleinste umschließende Rechteck (grün) sind für diesen Match ebenfalls eingezeichnet.