

Automatisierte Objektselektion und -verfolgung in einem adaptiven Farbraum

A. Al-Hamadi⁽¹⁾, H. Hofmeister⁽²⁾, B. Michaelis⁽¹⁾, B. Brückner⁽²⁾, R. Mecke⁽¹⁾
Institut für Elektronik, Signalverarbeitung und Kommunikationstechnik (IESK)⁽¹⁾
Otto-von-Guericke-Universität Magdeburg
Postfach 4120, D-39016 Magdeburg
Speziallabor Informatik, Leibniz-Institut für Neurobiologie⁽²⁾
Postfach 1860, D-39008 Magdeburg
Email: ayoub@iesk.et.uni-magdeburg.de, Email: hofmeister@ifn-magdeburg.de

1 Einleitung

Die Bewegtbildanalyse steht heute im Fokus vielfältiger Forschungen und erster Anwendungen. Für ungestörte Bildszenen und insbesondere für die Analyse von Grauwertbildern existieren leistungsfähige Verfahren für das Tracking von Objekten und die Bestimmung von Bewegungsvektorfeldern. Diese versagen jedoch häufig bei nichtkooperativen Messsituationen wie bspw. wechselnden Beleuchtungsverhältnissen oder Schattenwurf. Ein grundsätzliches Problem besteht auch in der automatisierten Festlegung des Startpunktes des Trackingalgorithmus und der Bestimmung der Blockgröße bei Verwendung von blockorientierten Matchingverfahren. Die Qualität und Geschwindigkeit der Objektverfolgung hängt im starken Maße vom berücksichtigten Vorwissen bzw. den beschränkenden Annahmen bzgl. des Objektes ab. Durch die Nutzung von Farbbildern sind generell qualitative Verbesserungen zu erwarten. Diese werden jedoch stark durch den verwendeten Farbraum beeinflusst.

2 Systemmodell

Im Beitrag wird ein Bildanalysesystem zur Selektion und Verfolgung bewegter Objekte in Farbbildsequenzen vorgestellt. Zur automatisierten initialen Objektselektion kommt eine pixelgenaue, bewegungsbasierte Farbklassifikation mit neuronalen Netzen zur Anwendung. So können beliebige, starre Objekte gefunden werden. Die Farbklassifikation ist gegenüber nichtkooperativen Messsituationen empfindlich. Deshalb wird für die anschließende Objektverfolgung die Verwendung eines adaptiven $K_1K_2K_3$ -Farbraums vorgeschlagen, der in Abhängigkeit von den Bildinhalten die genutzten Farbraumkomponenten wichtet und die Bestimmung des Verschiebungsvektors eines Objektes auch in ungünstigen Situationen gewährleistet [1]. Die Systemkonzeption ist in Abbildung 1 dargestellt. Die einzelnen Komponenten werden nachfolgend erläutert.

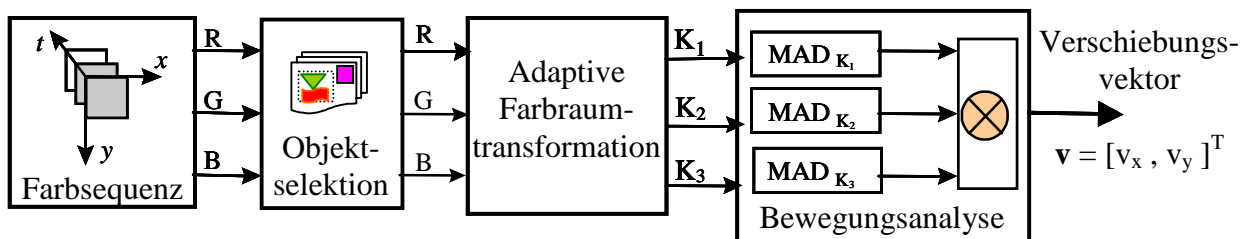


Abbildung 1: Vereinfachte Systemkonzeption

3 Objektselektion

Die erste Verarbeitungsstufe in der vorgeschlagenen Systemkonzeption ist die Objektselektion. Ziel des Bildanalyse-Systems ist die Verfolgung und Erkennung bewegter Objekte. Diese Objekte sollen möglichst wenigen Einschränkungen unterliegen, um die Analyse des Bildinhaltes nicht von vorhandenem Vorwissen abhängig zu machen. Die einzigen Voraussetzungen, die die Objekte erfüllen müssen, sind ihre Form- und Farbstabilität, sowie zur initialen Selektion die Bewegung.

Die Auswahl der Objektkandidaten erfolgt durch farbklassenbasiertes Blockmatching [2]. Dies ist ein neurobiologisch motivierter Ansatz, der grundlegende Prinzipien der Verarbeitung visueller Reize im Gehirn von Säugern auf ein technisches System überträgt. Grundlage neurobiologischer visueller Wahrnehmung ist insbesondere die Existenz paralleler Verarbeitungspfade für Bewegung und dynamische Form (magnozeller 'Wo-Pfad') bzw. Farbe und Form (parvozellulärer 'Was-Pfad'). Entlang dieser Verarbeitungspfade wird der visuelle Reiz rezeptiver Felder analysiert, in denen der Bildinhalt definierter Ausschnitte des Sehfeldes abgebildet ist. Die rezeptiven Felder vergrößern sich schrittweise auf dem Weg zur Erkennung komplexer Objekte in höheren visuellen Arealen. Im primären Sehfeld V1 existieren vollständige Analysemodule (Kolumnen) der rezeptiven Felder zur Erkennung von Signalform, Wellenlänge, Luminanz, Bewegung und Tiefe des visuellen Reizes.

Farbklassenbasiertes Blockmatching stellt einen einfachen Ansatz zur Übertragung der neurobiologischen Prinzipien der visuellen Wahrnehmung auf ein technisches System zur Bildanalyse dar. Der Ablauf der automatischen, farbbasierten Auswahl bewegter Objekte ist in Abbildung 2 schematisiert.

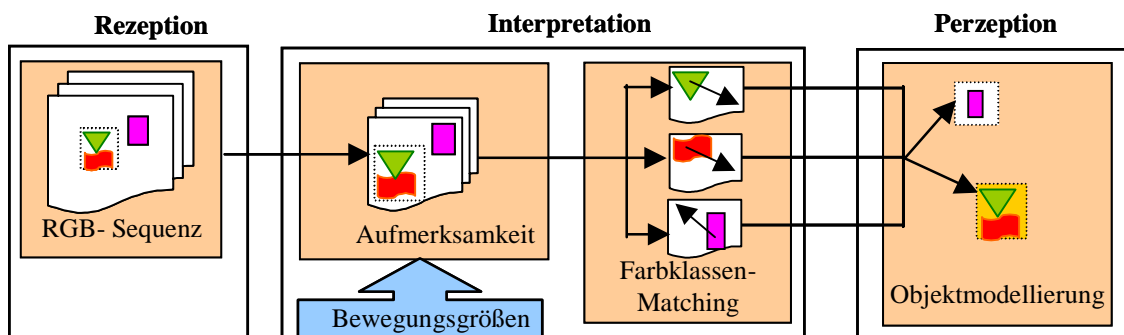


Abbildung 2: Objektselektion durch farbklassenbasiertes Blockmatching

Beim farbklassenbasierten Blockmatching werden die Bilder der im RGB-Farbraum aufgenommenen Szene in nichtüberlappende, quadratische Blöcke variabler Größe unterteilt. Die ersten Verarbeitungsschritte der Szenenanalyse - die Farbklassifikation und die Bestimmung der Bewegungsrichtung - sind lokal auf die Bildblöcke bezogen. Erst die Erkennung komplexer Objekte erfolgt global im Gesamtbild.

Für die Blöcke einer Szene erfolgt die pixelgenaue Farbklassifikation mittels eines unüberwacht lernenden künstlichen neuronalen Netzes. Konkret wird ein Fuzzy-ART-Netz [6] verwendet, das auf der Adaptiven Resonanztheorie von S. Grossberg basiert [5]. Der Klassifizierungsalgorithmus und Aspekte der Parametereinstellung der Fuzzy-ART-Netze sind in [7] erläutert.

Diese Netze klassifizieren beliebig dimensionale, positiv reellwertige Merkmalsvektoren entsprechend ihrer Lage im n -dimensionalen Merkmalsraum \mathbf{R}^n . Als Merkmale zur Farbklassifikation werden die 3-dimensionalen Farbvektoren der einzelnen Pixel verwendet. Der Merkmalsraum \mathbf{R}^{RGB} ist für alle Blöcke einer Szene gleich und die Klassifikationsergebnisse sind demzufolge vergleichbar.

Im Ergebnis der Klassifikation werden die Bildinhalte in binären Matrizen (siehe Abbildung 3) abgelegt, die unter Beibehaltung der lokalen und globalen Bildzusammenhänge die Farbklassenzugehörigkeiten aller Pixel eindeutig kodieren. Basierend auf diesen binären Matrizen wird der Blockmatchingalgorithmus für alle Farbklassen parallel angewandt.

In einem definierten Suchbereich eines Folgebildes der Szene wird der Bildinhalt auf Ähnlichkeit mit dem Inhalt des Referenzblockes des Ausgangsbildes geprüft, indem eine Maske in der Größe des Referenzblockes pixelweise im Suchbereich verschoben und ein Fehlerkriterium bestimmt wird. Bei maximaler Übereinstimmung bildet das Fehlerkriterium einen Extremwert.

Aufgrund des geringen Realisierungsaufwandes verwenden wir als Kriterium die mittlere absolute Differenz

$$\text{MAD}(v_x, v_y) = \frac{1}{255 \cdot M \cdot N} \sum_{y=y_0}^{y_0+N} \sum_{x=x_0}^{x_0+M} |I_k(x, y) - I_{k+1}(x + v_x, y + v_y)| \quad (1)$$

wobei M und N die Dimensionen des Referenzblockes I_k sind und v_x sowie v_y die Verschiebungen entlang der Bildachsen. Das Minimum stellt den Extremwert dar.

Im Ergebnis des Blockmatchings erhält man für jeden Referenzblock einen Verschiebungsvektor \mathbf{v} , der die geschätzte Bewegung des repräsentierten Bild- bzw. Objektausschnittes beschreibt. Die resultierenden Bewegungsvektorfelder der einzelnen Farbklassen des Gesamtbildes können überlagert werden. Dadurch wird ein Bildblock nicht mehr als Ganzes mit einem Vektor für die Bewegungsrichtung beschrieben, sondern stärker differenziert mit den Bewegungen für die einzelnen Farbklassen. Da Bildblöcke aufgrund ihrer starren Geometrie i.a. neben den interessierenden bewegten Objekten oder Objektausschnitten auch unbewegte Bildinhalte bzw. anders bewegte Objekte erfassen, lassen sich die gesuchten Objektkonturen anhand von Bildblöcken nur sehr grob und ungenau beschreiben. Die MAD-Funktion wird zudem durch irrelevante Bildbereiche innerhalb des Blockes unterschiedlich stark gestört. Durch farbklassenbasiertes Blockmatching ist die Bewegungsinformation mit der Farbinformation verbunden und lokale Bildzusammenhänge bleiben erhalten. Nur Blöcke gleichartiger Bewegung werden über die beteiligten Farbklassen zusammengefasst, so dass beliebig geformte Bildregionen aus Pixeln gleichartig bewegter Farbklassen selektiert werden können (siehe Abbildung 3). Diese Bildregionen stellen Objektkandidaten dar, deren Lage und Kontur das Trackingsystem initialisieren.

4 Adaptive Farbraumtransformation

Erfahrungsgemäß kann nur für Bildregionen mit ausgeprägten Intensitätsgradienten (z. B. Hell-Dunkel-Übergänge, Ecken, Texturen usw.) eine genaue und zuverlässige Bewegungsbestimmung realisiert werden. Deshalb wird als zweite Verarbeitungsstufe in der Systemkonzeption eine Farbraumtransformation zur Kontrasterhöhung vorgeschlagen, die adaptiv bezüglich des zu berechnenden Ähnlichkeitsmaßes zur Bestimmung der Bewegungsparameter ist.

Durch die folgende lineare Transformation erhält man für einen Bildpunkt im RGB-Farbraum die Komponenten K_i im $K_1K_2K_3$ -Farbraum:

$$\begin{bmatrix} K_1 \\ K_2 \\ K_3 \end{bmatrix} = [A] \cdot \begin{bmatrix} R \\ G \\ B \end{bmatrix} = \begin{bmatrix} q_{11} & q_{12} & q_{13} \\ q_{21} & q_{22} & q_{23} \\ q_{31} & q_{32} & q_{33} \end{bmatrix} \cdot \begin{bmatrix} R \\ G \\ B \end{bmatrix} \quad (2)$$

Die Transformationsmatrix $[A]$ dient hier als Wichtungsmatrix für die Umrechnung der Farbkanäle. Zur Bestimmung der Gewichte q_{ij} eignet sich die Karhunen-Loève-Transformation [4], bei der die Komponenten K_i im Sinne eines möglichst großen Kontrastes in Richtung größter Varianzen ausgerichtet werden (Hauptachsentransformation). Das entspricht einer Ermittlung der Eigenvektoren der Kovarianzmatrix der RGB-Bilddaten. Die orthonormalen Vektoren $\mathbf{q}_i = [q_{i1} \ q_{i2} \ q_{i3}]^T$ bilden die Achsen eines neuen Koordinatensystems. Der zum größten Eigenwert gehörige Eigenvektor $\mathbf{q}_1 = [q_{11} \ q_{12} \ q_{13}]^T$ stellt die Hauptachse dar und weist in Richtung größter Streuung der Farbwerte.

Die Transformation in den $K_1K_2K_3$ -Farbraum ähnelt in ihren Ergebnissen der Transformation in den HUV-Farbraum. Durch die jeweilige Adaption auf konkrete Bildregionen können sich aber Vorteile ergeben, insbesondere im Falle geringer Intensitätsgradienten. Die adaptive Farbraumtransformation weist gegenüber den mittelwertbasierten Transformationen (wie HUV) eine höhere Dynamik auf.

Die erste Komponente K_1 im adaptiven Farbraum ist achromatisch und beschreibt die Helligkeit. Für die betrachtete, objektangepasste Bildregion entsprechend Abbildung 4a enthält die Komponente K_1 Informationen, die etwa 90% der Gesamtvarianz der Bildregion entsprechen. Diese prozentuale Angabe ändert sich natürlich mit der jeweils betrachteten Region. Bewertungskriterium für die Transformation von Farbdaten in Grauwertdaten ist hierbei, wie man Primitiven bzw. Helligkeitsübergänge in allen drei Farbkomponenten möglichst ungeschwächt auf die Projektion (achromatische Komponente) überträgt. Bei Vergrößerung der Bildregion, wie in Abbildung 4b, gleicht sich die K_1 - der H-Komponente des HUV-Farbraums (Konvertierungsgerade zwischen Schwarzpunkt und Weißpunkt) immer stärker an. Bezüglich kleiner bzw. objektangepasster Bildregionen verstärkt K_1 die Primitiven, während sie bei der H-Komponente geschwächt werden.

Die K_2 - und K_3 -Komponenten im adaptiven Farbraum beschreiben Farbdifferenzen. Sie werden mit dem Ziel der Minimierung des Einflusses von Helligkeitsschwankungen und Schattenwurf ausgewertet.

Das Prinzip der Unempfindlichkeit von beiden helligkeitsunabhängigen Komponenten K_2 und K_3 gegenüber Schattenwurf und Beleuchtungsänderungen wird am Beispiel der betrachteten Bildregion aus Abbildung 4a im Folgenden erläutert.

Addiert bzw. subtrahiert man zu den Farbwerten in Gleichung 3a eine Konstante Δh als Maß für die Helligkeitsschwankung, dann zeigt sich, dass diese den Wert von K_3 innerhalb der Bildregion über die Sequenz kaum beeinflusst.

$$K_3 = 0.3237 \times R - 0.8134 \times G + 0.4833 \times B \approx 0.3237 \times (R + \Delta h) - 0.8134 \times (G + \Delta h) + 0.4833 \times (B + \Delta h) \quad (3a)$$

Aus der Gleichung 3a ergibt sich nämlich, dass die Summe der Faktoren näherungsweise Null (-0.0064) wird. Diese Gleichung gibt also die Beziehung wieder, die zwischen den beiden Komponenten $M=(R+B)$ und G in Bezug auf die betrachtete Bildregion in Abbildung 4a besteht.

$$K_2 = -0.741 \times R + 0.099 \times G + 0.663 \times B \approx -0.741 \times (R + \Delta h) + 0.099 \times (G + \Delta h) + 0.663 \times (B + \Delta h) \quad (3b)$$

Anhand der Gleichung 3b erkennt man die Beziehung, die zwischen den beiden Farbwerten R und B in Bezug auf die betrachtete Bildregion besteht. Die G -Komponente ist hier nicht dominant und vernachlässigbar.

Wie in den Gleichungen 3a und 3b zu sehen ist, ändern sich die K_2 - und K_3 -Komponenten nur wenig, wenn von jedem Farbwert eine Konstante Δh abgezogen bzw. addiert wird.

Dieses Ergebnis ist dem Verhalten der U- bzw. V-Komponente des HUV-Farbraumes ähnlich und es entspricht der visuellen Farbwahrnehmung - variiert die Sättigung einer Farbe, so bleibt der Buntton annähernd konstant.

5 Bewegungsanalyse

Nach der Transformation der Farbbilder in den adaptiven Farbraum werden diese in die dritte Verarbeitungsstufe der Systemkonzeption weitergeleitet. Dort wird die Bewegung der Bildregion (Bewegungsvektor $v = [v_x, v_y]^T$) über die Lagekoordinaten des MAD-Minimums ermittelt. Dieses Prinzip ist als ikonisches Matching-Verfahren [8] bekannt, wobei hier nicht feste Blöcke verwendet werden, sondern objektangepasste Bildregionen aus der ersten Verarbeitungsstufe der Systemkonzeption. Berechnet man nun die MAD-Funktion für alle theoretisch diskreten Bewegungsvektoren v_x und v_y in den Kanälen, so erhält man eine Matrix der Dimension: $(2 \times \Delta v_{\max} + 1) \times (2 \times \Delta v_{\max} + 1) \times 3$. Die Zahl 3 steht in dieser Matrixdimension für die Kanalanzahl.

Dieses Verfahren wird nachstehend als Multikanaliges Blockmatching (MKBM) bezeichnet. Der Vorteil des MKBM liegt im unterschiedlichen Verhalten der Kanäle. Beim Auftreten nichtkooperativer Messsituationen kann somit durch die gewichtete Kombination der drei Kanäle die Objektverfolgung verbessert werden. In der Abbildung 5 ist dies am Beispiel einer realen Bildsequenz dargestellt.

Die Problemstellung ist die Überlagerung eines bewegten Objektes (Auto) von einem Schattenwurf. Im Ergebnis dieser Überlagerung treten im Ähnlichkeitsmaß Deformationen und mehrere Minima auf. Dies geschieht sowohl im helligkeitsabhängigen Kanal K_1 sowie im Gesamtkriterium gemäß der Gleichung 1 wenn das Matching unter Verwendung der Bildinhalte in RGB-Farbraum durchgeführt wird. Daraus resultieren fehlerhafte Ergebnisse bei der Bestimmung der Bewegungsvektoren (siehe Abbildung 5a) und bei der Sequenzanalyse führt es dazu, dass die betrachtete Bildregion von der ursprünglich verfolgten Region wegdriftet (siehe Abbildung 5c). Die Ähnlichkeitsmaße der beiden Komponenten K_2 und K_3 ermöglichen aufgrund der guten Minimumausprägung hingegen eine fehlerfreie Bewegungsbestimmung.

Ausgehend von diesem Beispiel erscheint es sinnvoll, aus den kanalspezifischen Ähnlichkeitsmaßen des vorgeschlagenen adaptiven Farbraumes folgendes kombiniertes Gesamtkriterium abzuleiten:

$$MAD_{K_1 K_2 K_3}(v_x, v_y) = w_{1k} \cdot MAD_{K_{1k}}(v_x, v_y) + w_{2k} \cdot MAD_{K_{2k}}(v_x, v_y) + w_{3k} \cdot MAD_{K_{3k}}(v_x, v_y) \quad (5)$$

wobei K_{1k}, K_{2k}, K_{3k} : Kanäle im adaptiven Farbraum zum Zeitpunkt k

und w_{ik} : Wichtungsfaktoren mit $\sum_{i=1}^3 w_{ik} = 1$ zum Zeitpunkt k

Bei diesem Gesamtkriterium Gleichung 5 wird über die Faktoren w_i eine Wichtung der einzelnen kanalspezifischen Ähnlichkeitsmaße entsprechend ihrer Zuverlässigkeit ermöglicht. Untersuchungen zeigten, dass solche Wichtungsfaktoren mittels eines künstlichen neuronalen Netzes aus abgeleiteten Merkmalen des MAD-Kriteriums ermittelt werden können [3]. Die Bewegungsinformation unzuverlässiger Kanäle wird somit unterdrückt, wohingegen diejenigen Kanäle, die eine hohe Zuverlässigkeit erwarten lassen, mit einer hohen Wichtung einbezogen werden.

Das so ermittelte Gesamtkriterium für ein Beispiel (siehe Abbildung 5b) zeigt exemplarisch erreichbare Verbesserungen. In der Abbildung 5d sind die resultierenden Bewegungsvektoren in Form einer Bewegungstrajektorie im Vergleich zu dem herkömmlichen Verfahren (Abbildung 5c) dargestellt.

Bei der Verwendung des herkömmlichen Verfahrens beschreibt die Bewegungstrajektorie beim Eintritt des Objektes in den Schattenbereich nicht mehr die tatsächliche Objektbewegung. Dahingegen ermöglicht die vorgeschlagene Systemkonzeption die Verfolgung der betrachteten Bildregion über die gesamte Bildsequenz.

6 Zusammenfassung und Ausblick

In diesem Beitrag wurde ein System zur automatisierten Objektselektion und -verfolgung von Objekten in Farbbildsequenzen vorgeschlagen. Es wurde zur initialen Objektselektion eine pixelgenaue, bewegungsbasierte Farbklassifikation mit neuronalen Netzen angewendet. Sie ermöglicht die automatische Selektion von beliebig geformten Bildregionen, die Objektkandidaten repräsentieren. Da die Farbklassifikation im RGB-Farbraum gegenüber nichtkooperativen Messsituationen empfindlich ist, wurde die Einbeziehung eines adaptiven Farbraums bei der Bewegungsbestimmung vorgeschlagen. Es konnte gezeigt werden, dass durch eine adaptive Farbraumtransformation die Ähnlichkeitsmaßberechnungen von vornherein für Kanäle mit ausgeprägten Intensitätsgradienten realisiert werden können. Diese kanalspezifischen Ähnlichkeitsmaße werden entsprechend ihrer Zuverlässigkeit zu einem kombinierten Gesamtkriterium zusammengefasst, das insbesondere in nichtkooperativen Messsituationen durch adaptive Wichtung der Anteile genauere und robustere Ergebnisse, verglichen mit herkömmlichen Verfahren, liefert. Die Wichtung erfolgt durch Auswertung von Merkmalen der Ähnlichkeitsmaße.

Danksagung

Diese Arbeit wird durch das Land Sachsen-Anhalt gefördert (FKZ: 1441A/8386H und 1441B/8386H).

Literatur

- [1] Al-Hamadi, A., Mecke, R., Michaelis, B.: Verbesserte Bewegungsbestimmung bei der adaptiven Farbraumtransformation. 5. Workshop Farbbildverarbeitung, 7.-8.10.99, TU-Ilmenau, 1999
- [2] Hofmeister, H., Brückner, B.: Color Classification to improve Block-Based Motion Estimation in RGB-Image Sequences. Proceedings of 6th ICONIP 1999, Vol.3, pp.1224-1229, 1999
- [3] Mecke, R.: Grauwertbasierte Bewegungsschätzung in monokularen Bildsequenzen unter besonderer Berücksichtigung bildspezifischer Störungen. Dissertation, Uni-Magdeburg, 1999
- [4] Wahl, F. M.: Digitale Bildsignalverarbeitung, Springer-Verlag 1989.
- [5] Grossberg, S.: Adaptive pattern classification and universal recoding, II: Feedback, Expectation, Olfaction, Illusions. Biological Cybernetics, Vol. 23, pp. 187-202, 1976
- [6] Carpenter, G. A.; Grossberg, S.; Rosen, D.B.: Fuzzy-ART: Fast stable learning and categorization of analog patterns by an Adaptive Resonance System. Neural Networks, Vol. 4, pp. 759-771, 1991
- [7] Hofmeister, H.; Brückner, B.: Aspekte der Anwendung von Fuzzy-ART-Netzen zur Farbsegmentierung. 4. Farbworkshop Farbbildverarbeitung, ed. Rehrmann, V., Koblenzer Schriften zur Informatik, Vol. 9, pp. 25-32, Verlag Dietmar Fölbach, 1998
- [8] Rehrmann, V.: Korrespondenzanalyse in Farbbildern, Dynamische Perzeption, eds. Posch, St., Ritter, H., Proceedings in artificial intelligence, Vol. 8, S. 23-30, Infix, 1998

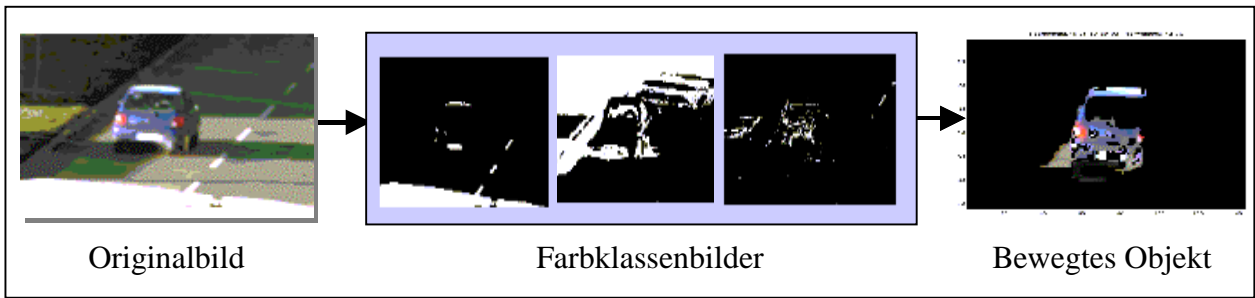


Abbildung 3: Farbklassenbasiertes Blockmatching zur Objektselektion

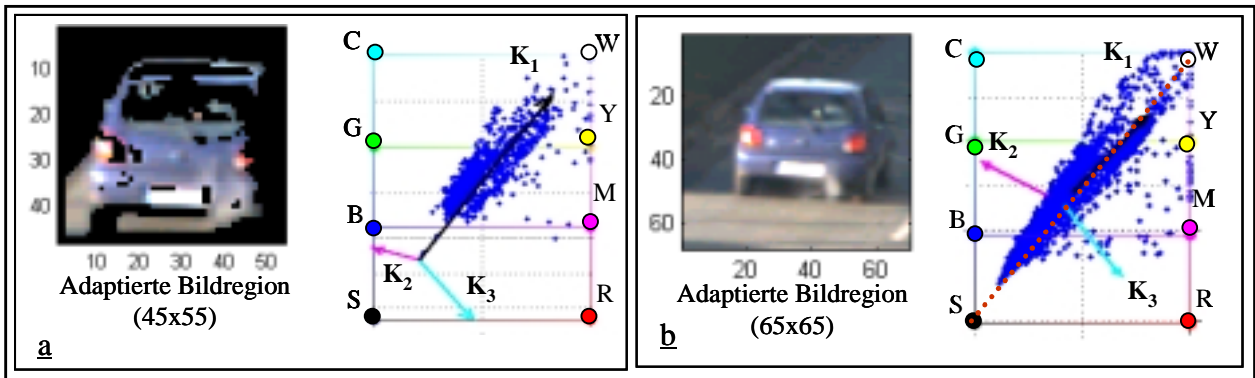


Abbildung 4: Adaptive Farbraumkomponenten für die jeweiligen Bildregionen

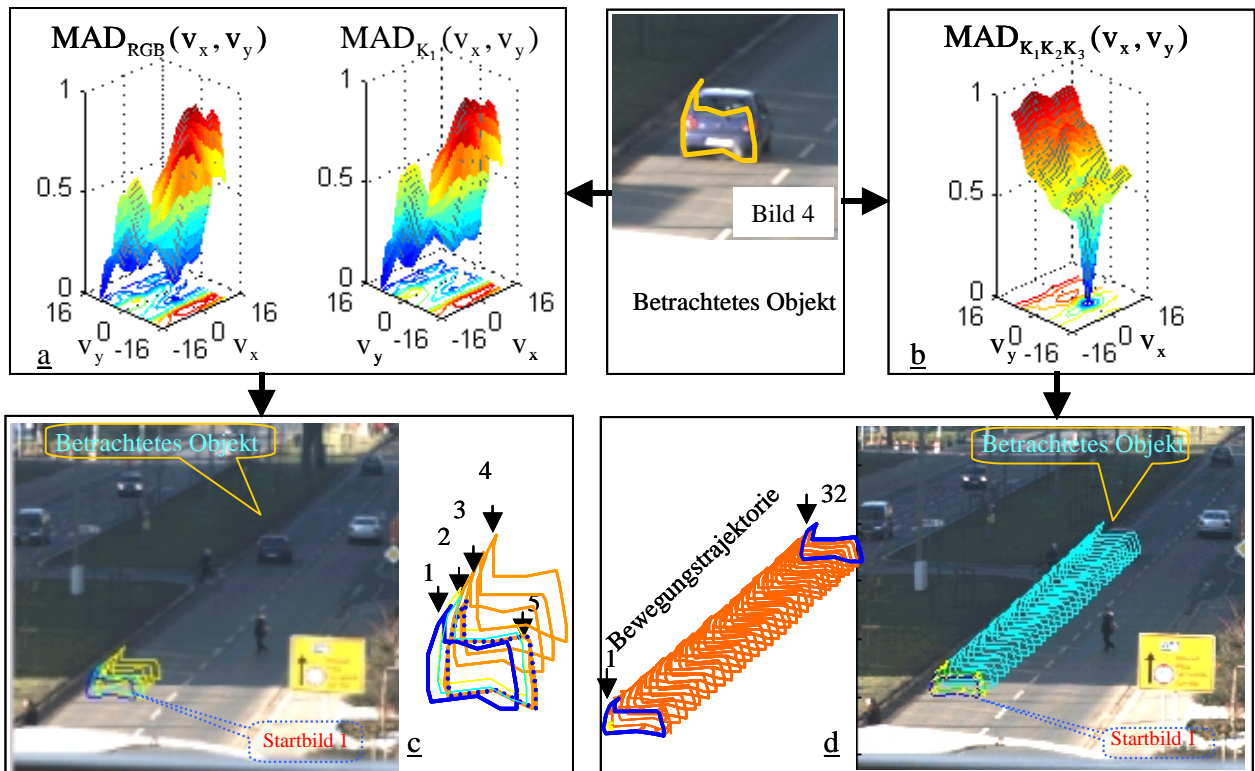


Abbildung 5: Ausschnitte aus einer realen Bildsequenz und verschiedene Ähnlichkeitsmaße a und b für eine exemplarische Bildregion bei Schattenwurf. Unten sieht man die Analyseergebnisse als Bewegungstrajektorie bei herkömmlichen Verfahren c und gemäß der Systemkonzeption d.