

11. Analyse von Bewegtbildern

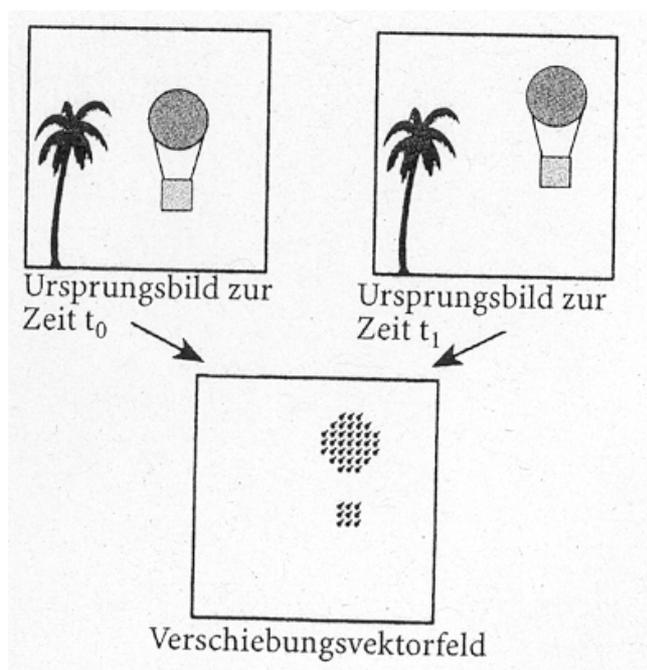
Problemstellung:

Analyse und Verstehen von Bewegung in Bildfolgen

Schwierigkeiten:

- Unterschiede zwischen Bildveränderung und Bewegung
- Rauschen soll nicht als Bewegung interpretiert werden
- ebenso Veränderung der Beleuchtungsbedingungen
- bewegte Teile eines Bildes müssen nicht notw. zusammenhängen (Beisp. Vogelschwarm) – Mensch erkennt dies trotzdem als zusammengehörig (Gestaltgesetz des "gemeinsamen Schicksals")
- Bewegung eines Objektes hinter einem Lattenzaun: ständig tauchen Bildelemente auf und verschwinden wieder
- Verformung von Objekten während der Bewegung: Beisp. Gehen eines Menschen, Aufprallen eines Balls, sich drehender Stein im Flug (nach Wurf) – auch Beleuchtung kann sich dabei ändern
- Eigenbewegung zu unterscheiden von Bewegung der beobachteten Objekte

Ziel der Bewegungsschätzung: Verschiebungsvektor für jeden Bildpunkt



Bewegungsdetektion

Annahme: Kamera statisch (Überwachungsaufgaben)

Ziel: Detektion von Objekten in Bewegung gegenüber einem konstanten Hintergrund

Variante 1:

$$d(r, c) = \begin{cases} 1, & \text{falls } |f(r, c, t) - f(r, c, t - 1)| \geq \epsilon \\ 0, & \text{sonst} \end{cases}$$

Variante 2:

$$d(r, c) = \begin{cases} 1, & \text{falls } |f(r, c, t) - f^r(r, c)| \geq \epsilon \\ 0, & \text{sonst} \end{cases}$$

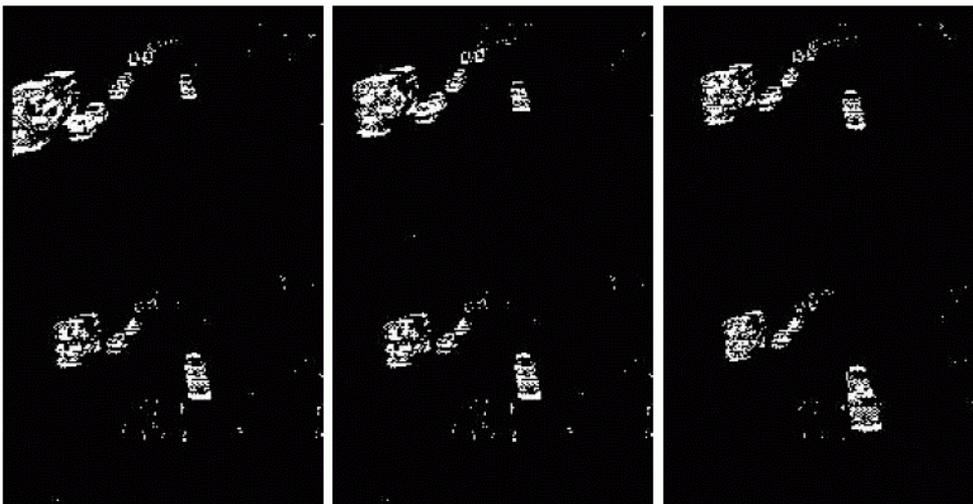
wobei $f^r(r, c)$ ein Referenzbild nur mit dem Hintergrund

Beispiel: Ausgangs-Bildfolge:

Bildsequenz: ≈ 5 Bilder/s

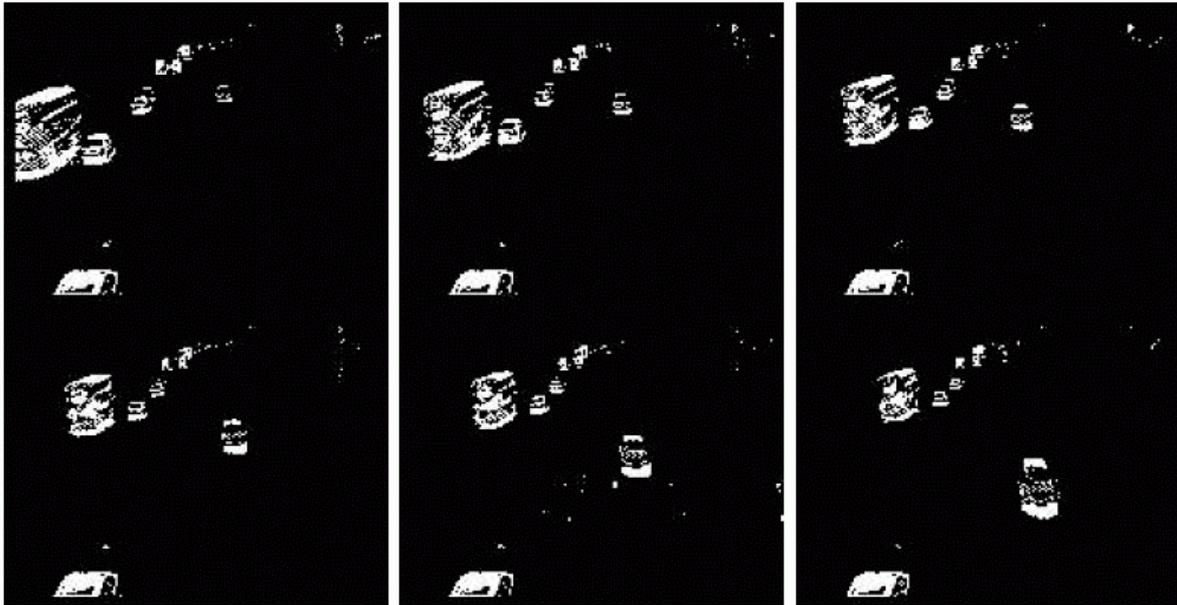


Bildsubtraktion: Variante 1



Schwäche: Doppelbild eines Fahrzeugs (aus dem letzten und aktuellen Bild); Aufteilung einer konstanten Fläche

Bildsubtraktion: Variante 2



Generell: weitere Nachverarbeitungsschritte nötig, um die Objekte zu lokalisieren (Zusammenhangsanalyse, Entfernen kleiner Regionen, Fusion benachbarter Regionen, Berechnung von Bounding-Box)

Schwelwert ϵ :

Aufnahme einer Reihe von Bildern nur mit dem Hintergrund. Histogramm aller Differenzwerte; sie sind auf Zufallsstörungen zurückzuführen und lassen sich als eine Gauss-Funktion modellieren. Der Schwellwert ϵ wird auf ein Mehrfaches der entsprechenden Standardabweichung gesetzt.

Referenzbild $f^r(r, c)$:

Mittelung einer langen Sequenz von Bildern

(Jiang 2006)

2 Ansätze zur Abtastung bei Bewegungsdetektion:

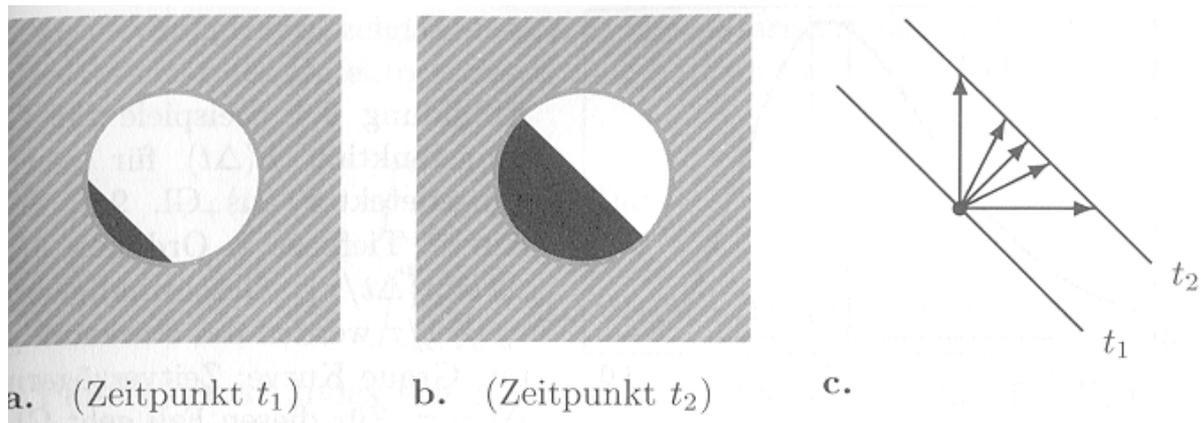
- zeitdiskret: Abfolge von Einzelbildern zu verschiedenen Zeitpunkten ("frames") – abgetastete Bewegung (*apparent motion*)
- kontinuierliche Entwicklung des Grauwertes an einem festen Ort: entspricht eher dem biologischen Vorbild (rezeptive Felder der retinalen Ganglienzellen; Insektenauge)

im zeitdiskreten Fall:

Bildelemente müssen von einem Bild zum nächsten verfolgt werden – *Korrespondenzproblem*

Problem beim 2. Ansatz (streng lokaler Detektor):

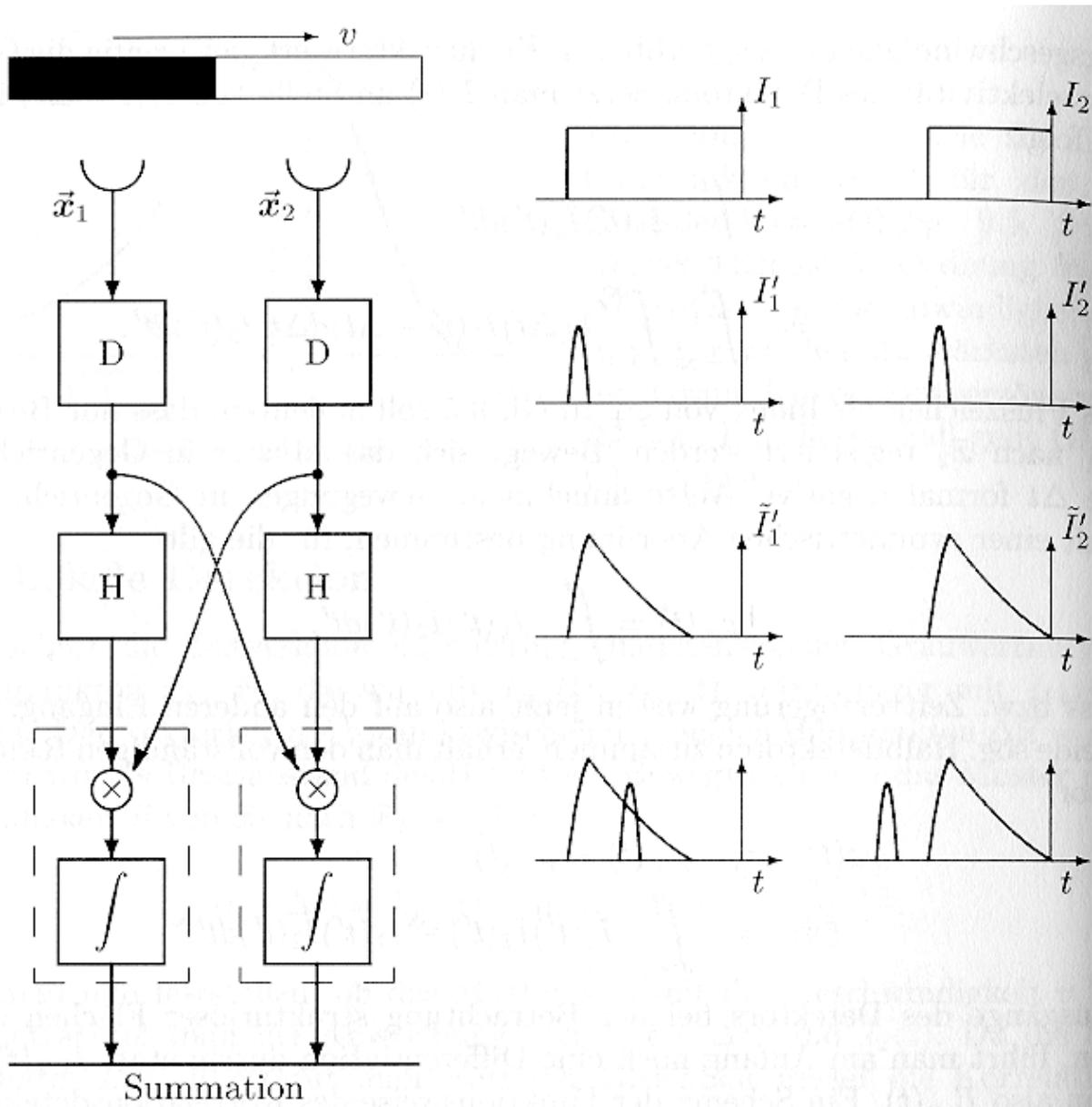
Helligkeitsänderungen an einem Punkt können auf ganz verschiedene Bewegungen zurückzuführen sein – *Aperturproblem*



(aus Mallot 2000)

Korrelationsdetektor nach Hassenstein & Reichardt (in Anlehnung an das Bewegungssehen von Insekten):

- (nahezu) kontinuierliche Verfolgung der Bildintensitäten an zwei Bildpunkten
- durch Vergleich der Zeitverläufe der Grauwerte wird auf Verschiebungen des Musters geschlossen
- verwendet (zeitliche) Änderungsdetektion (Ableitung nach der Zeit, D), Tiefpassfilter (H), Produktbildung und Integration



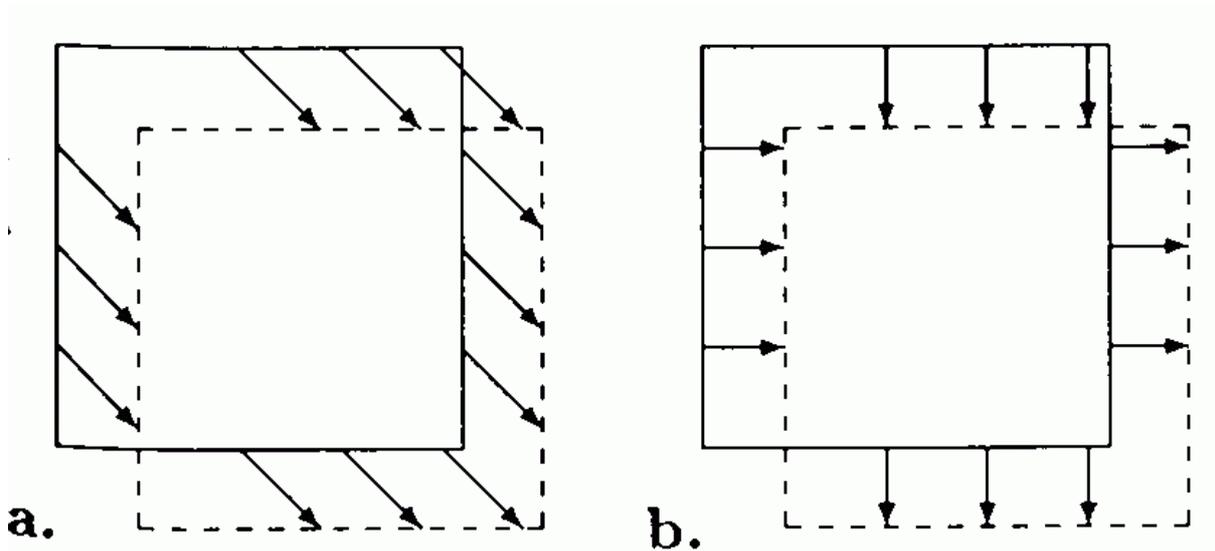
Gradientenverfahren messen für jeden Bildpunkt die zeitliche Änderung des Grauwertes
 Aperturproblem: nur die Bewegungskomponente senkrecht zur Kontur (d.h. in Richtung des Grauwertgradienten) wird dabei erfasst
 Hypothese: Bewegung findet (lokal) tatsächlich in diese Richtung statt:

$$\vec{v} = -\frac{\partial I(x, y, t)}{\partial t} \frac{\text{grad } I(x, y)}{\|\text{grad } I(x, y)\|^2}$$

Beispiel: diagonale Bewegung eines Quadrats

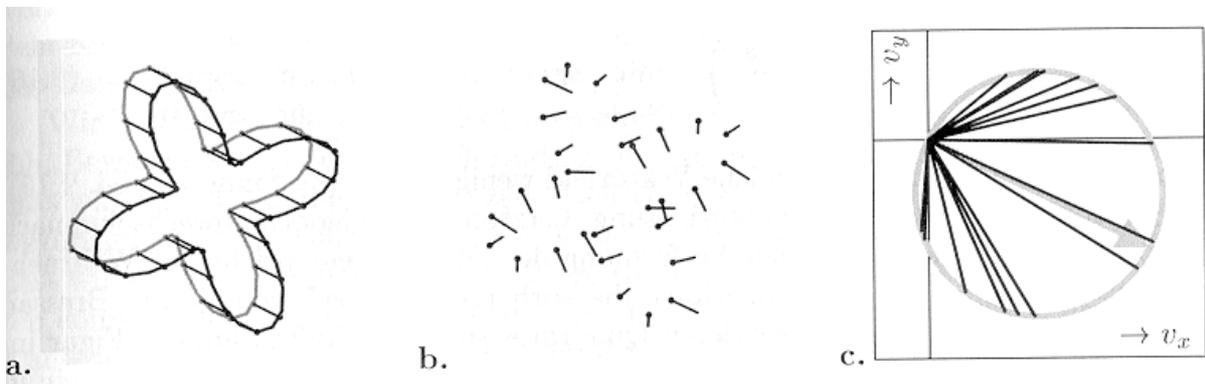
a. tatsächliche Bewegungsvektoren

b. nach obiger Gleichung errechnete Bewegungsvektoren



Bei konstantem Verschiebungsfeld (Bewegung = reine Translation) kann der tatsächliche Verschiebungsvektor dennoch aus diesen Daten geschätzt werden:

Ähnlich zur modifizierten Hough-Transformation von Geraden durch einen Fluchtpunkt (s. Übung, Aufg. U26) liegen alle Endpunkte von errechneten Verschiebungsvektoren auf einem Thales-Kreis durch den Nullpunkt. Der am Nullpunkt ansetzende Durchmesser dieses Kreises entspricht der tatsächlichen Verschiebung:



Optischer Fluss

Bewegungsfeld: Projektion der Bewegungen der 3D-Szene auf die Bildebene. Diese Größe soll aus einer Bildsequenz extrahiert werden.

Optischer Fluss: beobachteter Grauwertfluss in der Bildebene (Objektbewegung bewirkt, dass Grauwerte über die Bildebene fließen)

Annahme: Optischer Fluss \equiv Bewegungsfeld

Gegenbeispiel:

- Bewegungsfeld $\neq 0$, optischer Fluss = 0
 - Fläche konstanter Helligkeit
 - Eine sich drehende Kugel mit einer beliebigen gleichmäßigen Oberfläche. Eine solche Kugel kann sich durch ihren Schwerpunkt um jede Achse drehen, ohne dass Grauwertveränderungen resultieren.
- Bewegungsfeld = 0, optischer Fluss $\neq 0$

Die gleiche Kugel in Ruhe und beleuchtet von einer sich bewegenden Lichtquelle.

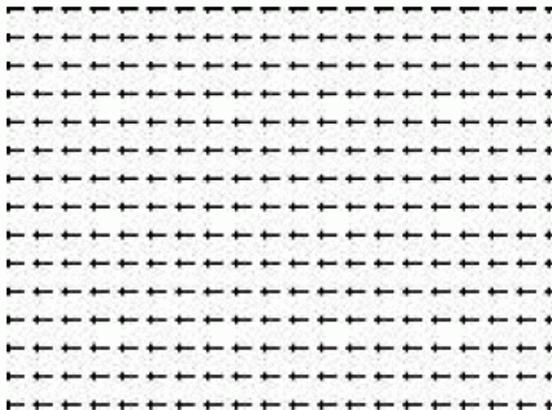
Obwohl diese Äquivalenz (Optischer Fluss \equiv Bewegungsfeld) zunächst plausibel erscheint, gilt sie bei genauerer Betrachtung tatsächlich in nur wenigen Fällen. In praktischen Anwendungen wird jedoch von einer ungefähren Äquivalenz ausgegangen.

Annahmen zur Berechnung des optischen Flusses:

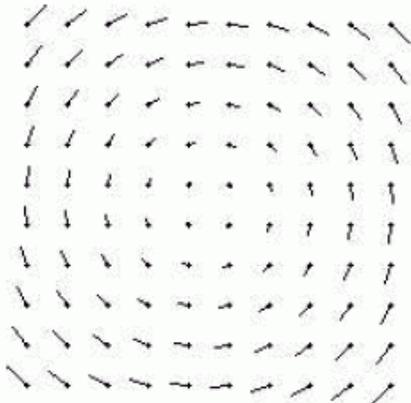
Der Grauwert eines Objektpunktes bleibt ungefähr unverändert

- keine (gravierende) Änderung der Beleuchtung von Bild zu Bild
- keine (gravierende) Änderung der Entfernung der Kamera und Lichtquelle(n) zu Objekten von Bild zu Bild

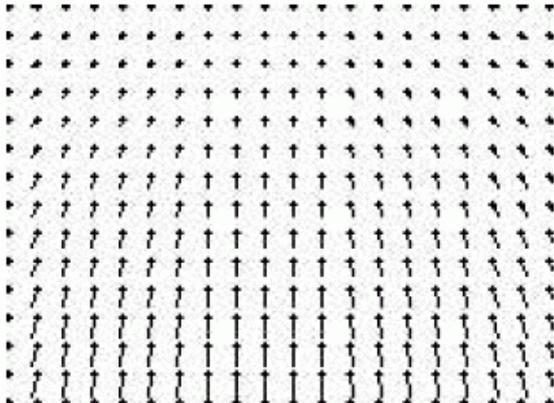
Beispiele für optischen Fluss:



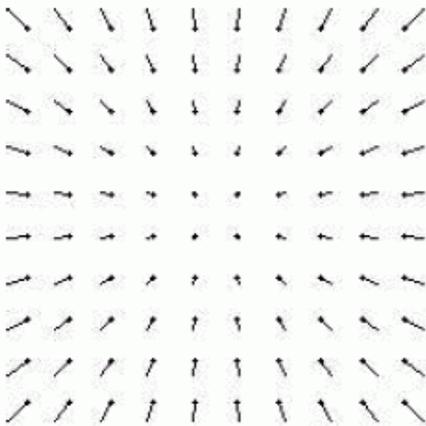
Translation



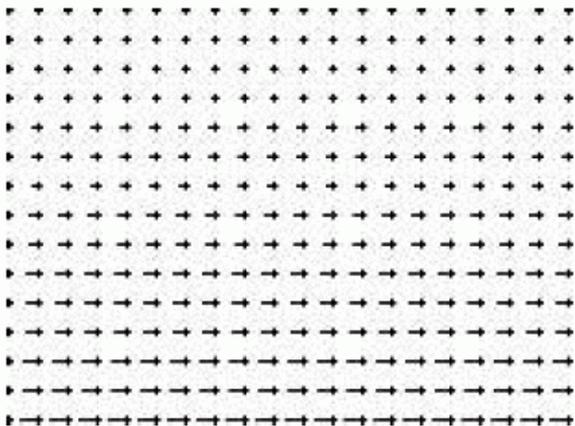
Rotation



Vorwärtsbewegung auf einer Ebene

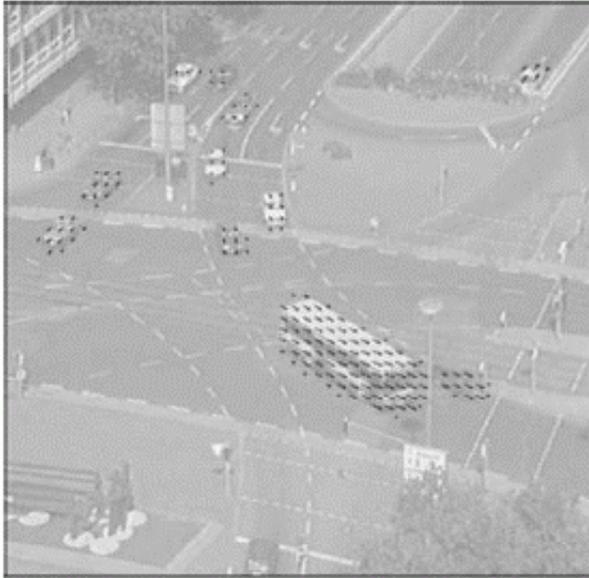


Zoom-in

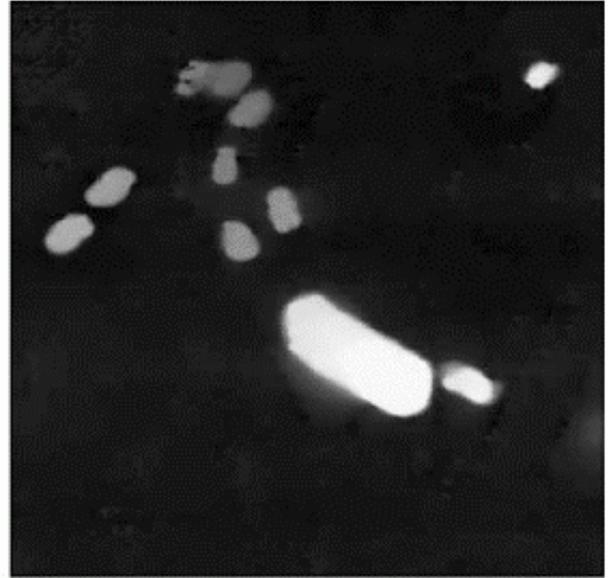


Seitwärtsbewegung

optischer Fluss in Realszenen (Verkehrsanalyse):



Optische Flüsse



Betrag der Flussvektoren

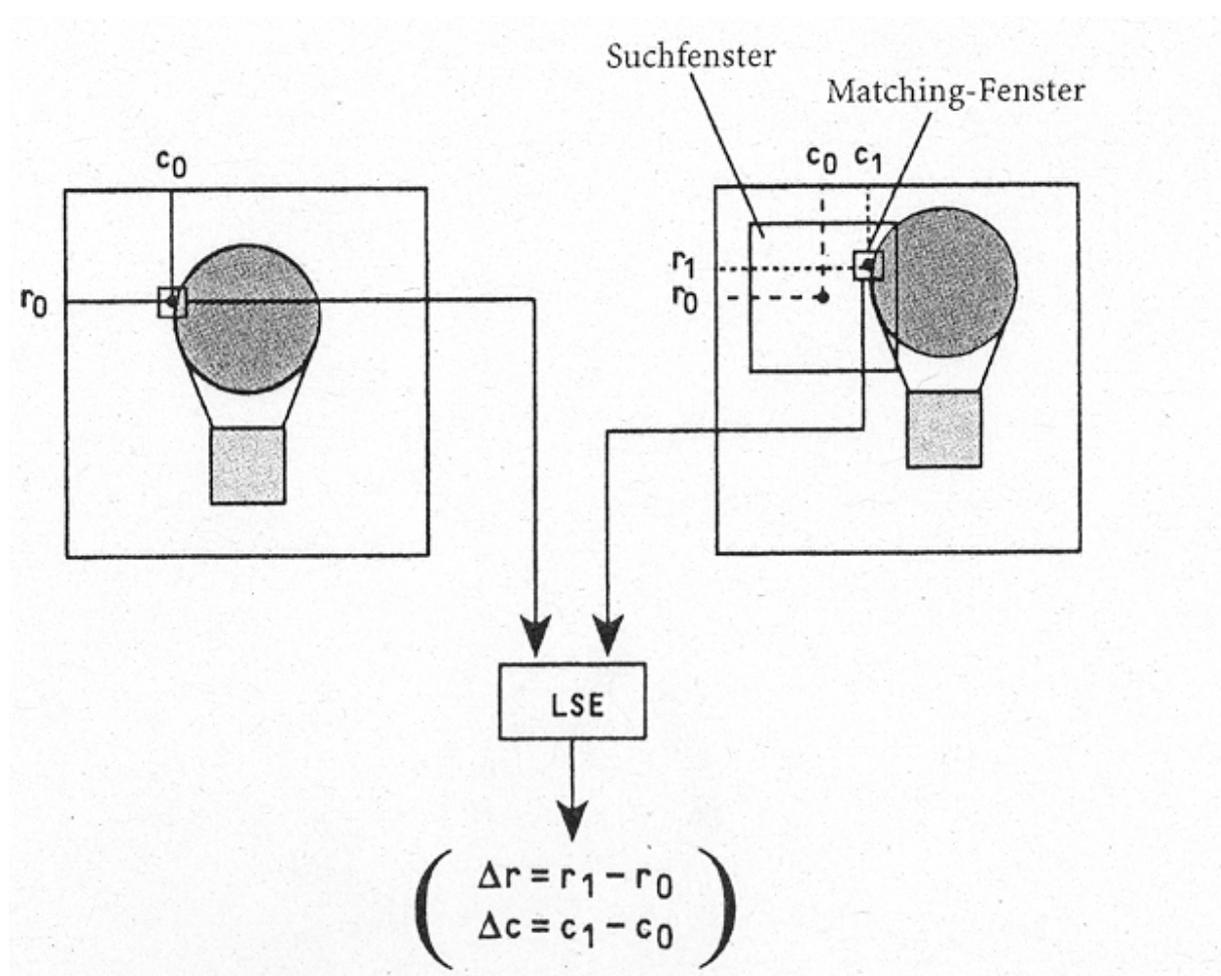
(aus Jiang 2006)

Bestimmung optischer Flüsse:

- lokale Verfahren
- globale Verfahren

Korrelationsverfahren zur lokalen Bewegungskdetektion

- lege Matching-Fenster in Bild 1 um aktuelles Pixel
- suche mittels Matching-Fenster der gleichen Größe in Bild 2 Übereinstimmung innerhalb eines Suchfensters
- für jedes Pixel im Suchfenster werden die Grauwerte der beiden Matching-Fenster verglichen
- der Vergleich, der den kleinsten quadratischen Fehler (LSE) aufweist, liefert die Verschiebungsdaten
- nachfolgende Korrekturverfahren zum Ausgleich von Korrespondenz- und Aperturfehlern



klassisches Verfahren zur Bewegungsdetektion
(beinhaltet initiale Verschiebungsdetektion und Korrektur):

Verfahren von Horn und Schunk

- iteratives Verfahren
- wird auf das gesamte Bild E (zum Zeitpunkt t_0 und t_1) angewandt
- 1. Schritt: Berechnung der partiellen Ableitungen

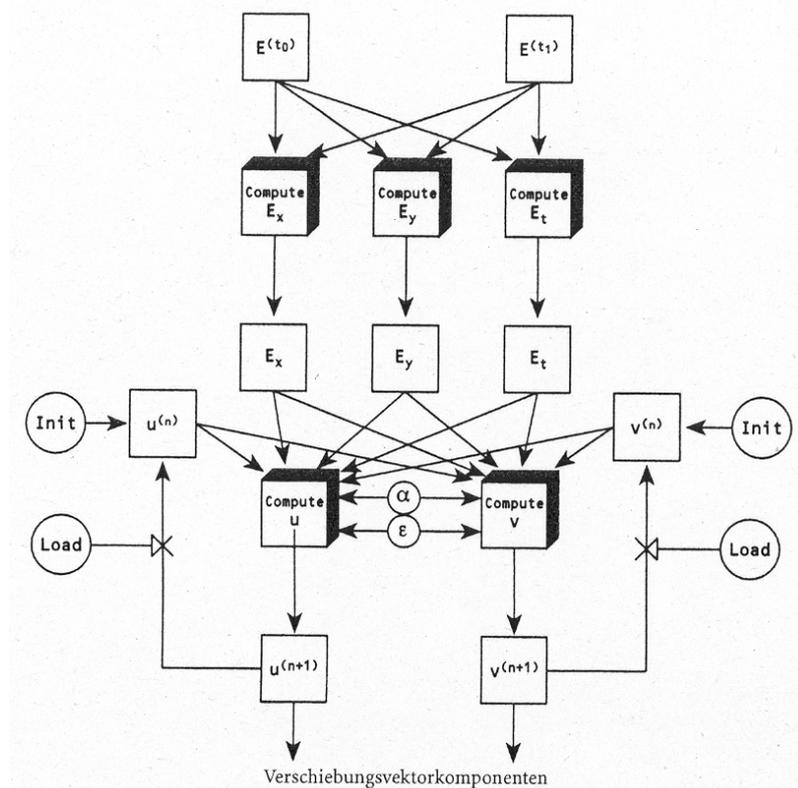
$$E_x = \frac{\partial E}{\partial x} \quad E_y = \frac{\partial E}{\partial y} \quad E_t = \frac{\partial E}{\partial t}$$

- Ziel: Komponenten u , v der Verschiebungsvektoren (Vektorfeld)

$$u = \frac{dx}{dt} \quad v = \frac{dy}{dt}$$

- werden nicht exakt bestimmt, sondern approximativ mit gegebener Fehlerschranke

Ablaufschema des Verfahrens:



Herleitung des Verfahrens:

$E(x, y, t)$ = Grauwert eines Pixels zum Zeitpunkt t

zum nächsten Zeitpunkt $t + \delta t$:

$$E(x, y, t) = E(x + \delta x, y + \delta y, t + \delta t)$$

Taylor-Entwicklung des rechten Terms um den Punkt (x, y, t) :

$$E(x, y, t) = E(x, y, t) + \delta x \frac{\partial E}{\partial x} + \delta y \frac{\partial E}{\partial y} + \delta t \frac{\partial E}{\partial t} + R$$

Es folgt:

$$\delta x \frac{\partial E}{\partial x} + \delta y \frac{\partial E}{\partial y} + \delta t \frac{\partial E}{\partial t} + R = 0$$

Vernachlässigung des Restglieds R und Division durch δt liefert:

$$\frac{\delta x}{\delta t} \frac{\partial E}{\partial x} + \frac{\delta y}{\delta t} \frac{\partial E}{\partial y} + \frac{\partial E}{\partial t} = 0$$

\Rightarrow im Grenzübergang $\delta t \rightarrow 0$:

$$\frac{\partial E}{\partial x} \frac{dx}{dt} + \frac{\partial E}{\partial y} \frac{dy}{dt} + \frac{\partial E}{\partial t} = 0$$

oder kurz

$$E_x u + E_y v + E_t = 0$$

Dies ist eine Gleichung mit 2 Unbekannten (u, v).

2. Bestimmungsgleichung: "Smoothness Constraint"; Annahme: benachbarte Punkte vollführen ähnliche Bewegungen ("glatte Bewegung").

Zur formalen Beschreibung dieser Bedingung wird gefordert, dass die beiden Terme

$$\left(\frac{\partial u}{\partial x}\right)^2 + \left(\frac{\partial u}{\partial y}\right)^2 \quad \text{und} \quad \left(\frac{\partial v}{\partial x}\right)^2 + \left(\frac{\partial v}{\partial y}\right)^2$$

bzw. ihre Summe = 0 werden soll.

Anstatt Gleichheit in beiden Bedingungen werden 2 Fehlerterme definiert, die die Abweichung von der Gleichheit quantifizieren:

$$\varepsilon_b = E_x u + E_y v + E_t$$

$$\varepsilon_c^2 = \left(\frac{\partial u}{\partial x}\right)^2 + \left(\frac{\partial u}{\partial y}\right)^2 + \left(\frac{\partial v}{\partial x}\right)^2 + \left(\frac{\partial v}{\partial y}\right)^2$$

Diese Fehler werden für jedes Pixel der Eingabebilder ermittelt; der Gesamtfehler

$$\varepsilon^2 = \iint (\varepsilon_b^2 + \alpha^2 \varepsilon_c^2) dx dy$$

soll minimiert werden.

Dabei steuert der Parameter α das Verhältnis des Einflusses der Einzelfehler auf den Gesamtfehler.

Abhängigkeiten der zu minimierenden Funktion:

$$\iint F(x, y, u, v, u_x, u_y, v_x, v_y) dx dy$$

Approximative Lösung des Extremalproblems mit Mitteln der Variationsrechnung (kurze Einführung bei Bässmann & Kreyss 1998, S. 407 ff.):

2 Euler-Gleichungen

$$\frac{\partial F}{\partial u} - \frac{\partial}{\partial x} \left(\frac{\partial F}{\partial u_x} \right) - \frac{\partial}{\partial y} \left(\frac{\partial F}{\partial u_y} \right) = 0$$

$$\frac{\partial F}{\partial v} - \frac{\partial}{\partial x} \left(\frac{\partial F}{\partial v_x} \right) - \frac{\partial}{\partial y} \left(\frac{\partial F}{\partial v_y} \right) = 0$$

werden approximativ umgeformt zu

$$\left(\alpha^2 + E_x^2 \right) u + E_x E_y v = \alpha^2 \bar{u} - E_x E_t$$

$$E_x E_y u + \left(\alpha^2 + E_x^2 \right) v = \alpha^2 \bar{v} - E_y E_t$$

mit lokalen Mittelwerten \bar{u} , \bar{v} .

Iterative Lösung des linearen Gleichungssystems mit dem Gauß-Seidel-Verfahren möglich:

$$u^{(n+1)} = \frac{\alpha^2 \bar{u}^{(n)} - E_x E_t - E_x E_y v^{(n)}}{\alpha^2 + E_x^2}$$

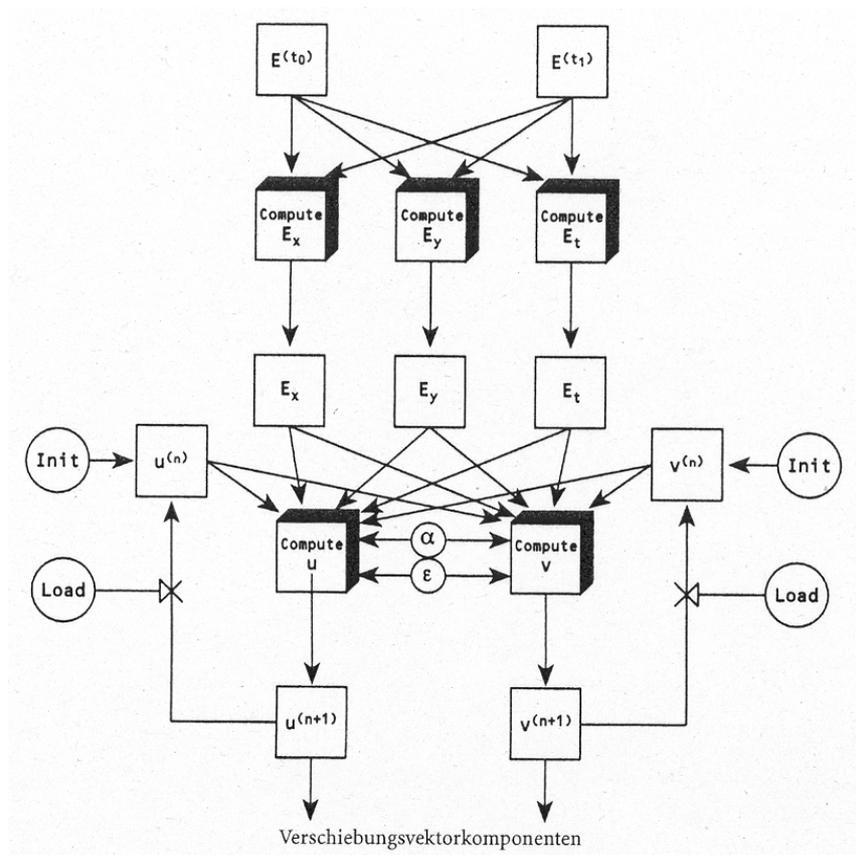
$$v^{(n+1)} = \frac{\alpha^2 \bar{v}^{(n)} - E_y E_t - E_x E_y u^{(n)}}{\alpha^2 + E_y^2}$$

kann noch so umgestellt werden, dass u und v isoliert auftreten (schnellere Berechenbarkeit, da Teile der rechten Seiten in beiden Iterationsgleichungen identisch sind):

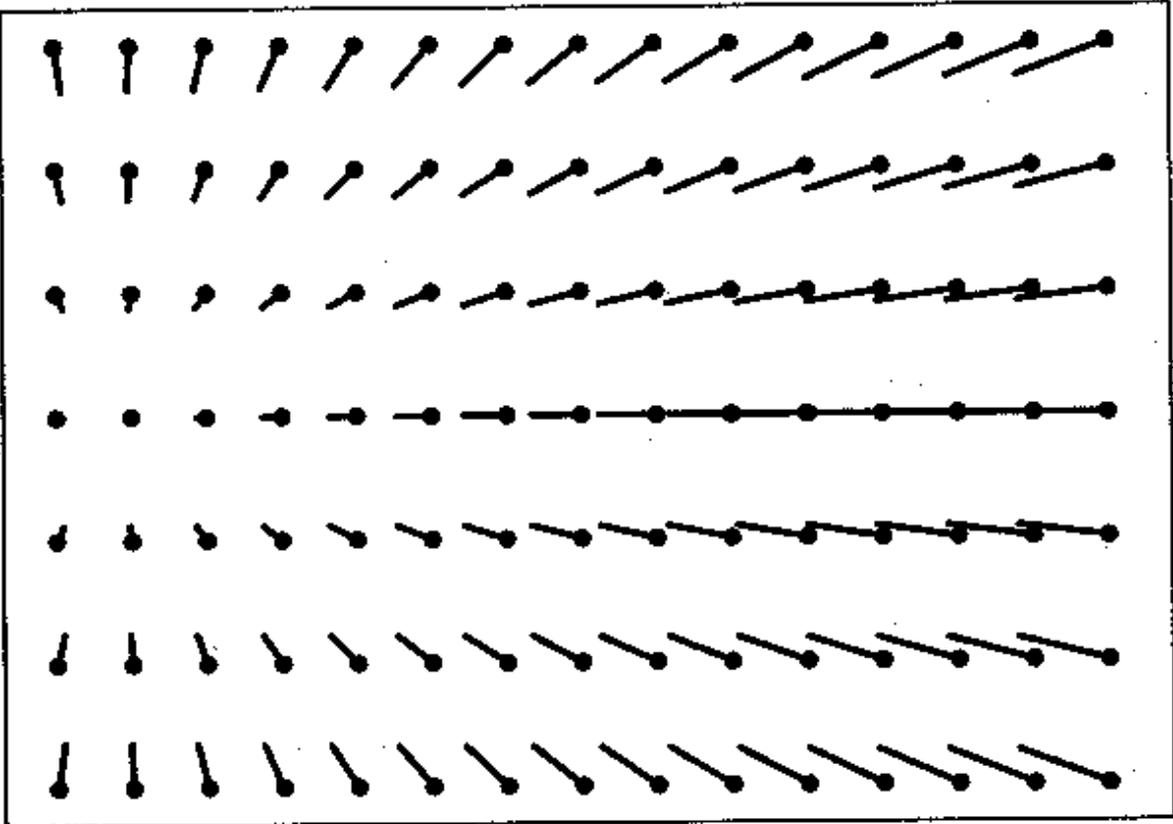
$$u^{(n+1)} = \bar{u}^{(n)} - \frac{E_x \left(E_x \bar{u}^{(n)} + E_y \bar{v}^{(n)} + E_t \right)}{\alpha^2 + E_x^2 + E_y^2}$$

$$v^{(n+1)} = \bar{v}^{(n)} - \frac{E_y \left(E_x \bar{u}^{(n)} + E_y \bar{v}^{(n)} + E_t \right)}{\alpha^2 + E_x^2 + E_y^2}$$

Iteration, bis Differenz zwischen aufeinanderfolgenden Werten vorgegebene Genauigkeitsschranke unterschreitet.



Ergebnis: "Nadelbild"



Punktverfolgung in Bewegtbildern

Ziel: Anstatt eines dichten Feldes des optischen Flusses sollen lediglich einzelne Punkte über eine Bildsequenz verfolgt werden

Prinzip: Im ersten Bild sollen die Punkte (points of interest) manuell/automatisch bestimmt werden. Ihre Positionen in den Folgebildern werden mittels Template-Matching in einer Nachbarschaft gesucht.

Template $t(i, j)$: Bildbereich um einen zu verfolgenden Punkt

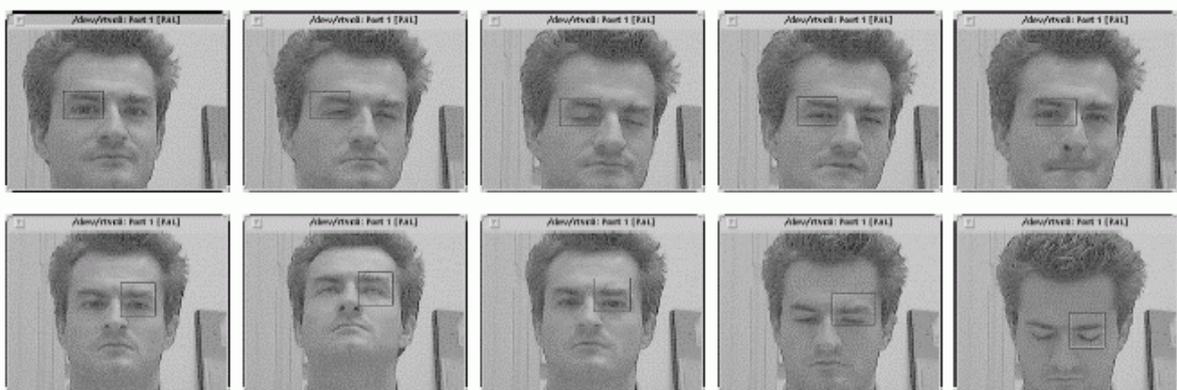
Suchbereich: Nachbarschaft um den zu verfolgenden Punkt im Folgebild

Ähnlichkeitsmaß:

- $d_1(t, g) = \sum_{i=1}^m \sum_{j=1}^n |g(i, j) - t(i, j)|$ (möglichst klein)
- $d_2(t, g) = \sum_{i=1}^m \sum_{j=1}^n (g(i, j) - t(i, j))^2$ (möglichst klein)
- $d_3(t, g) = \sum_{i=1}^m \sum_{j=1}^n g(i, j)t(i, j)$ (möglichst groß)

Bemerkung: Verbesserung durch Kalman-Filter im Rahmen der optimalen Schätztheorie

Beispiel: Augenverfolgung



(aus Jiang 2006)