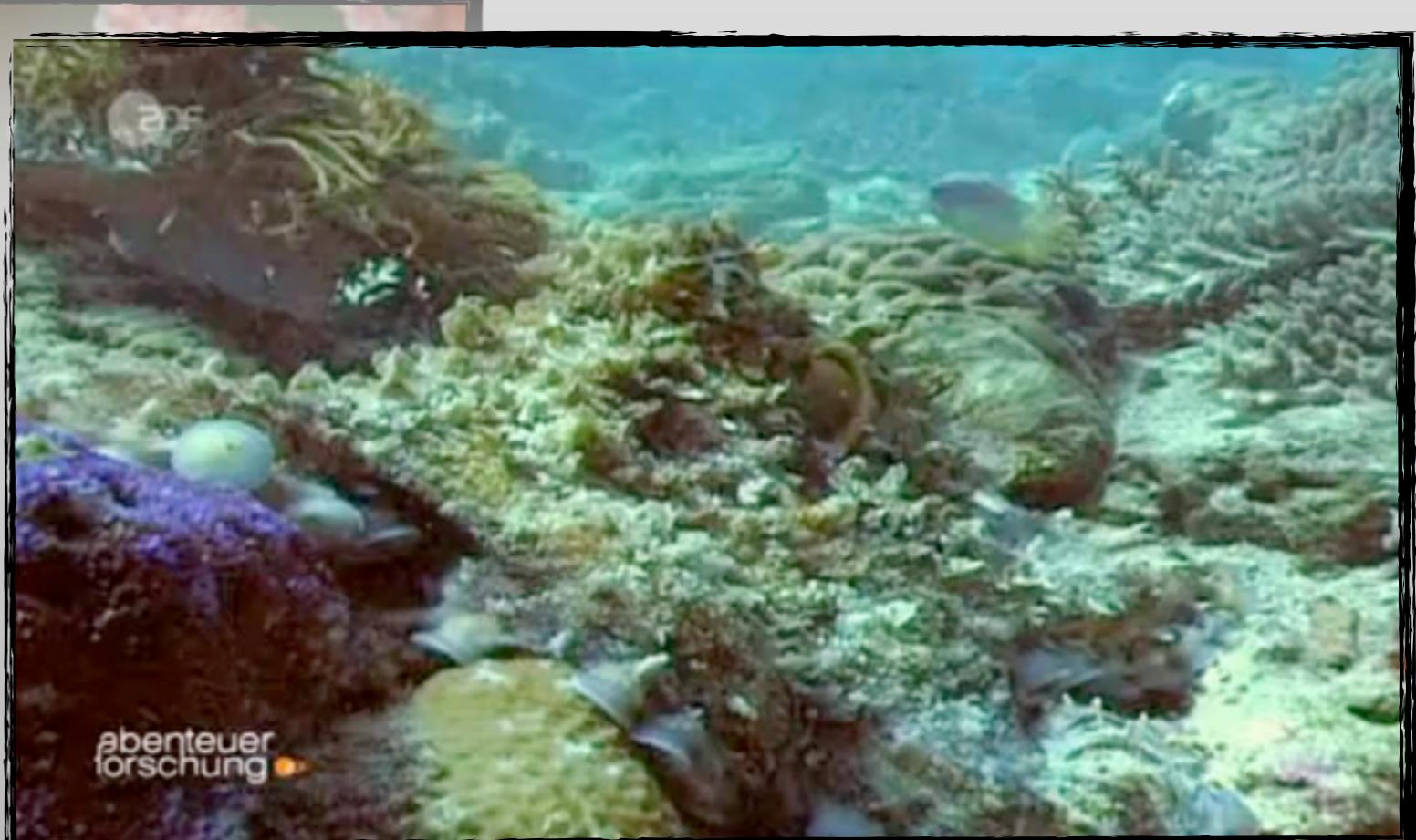


Object Segmentation in Video

A Hierarchical Variational Approach for Turning
Point Trajectories into Dense Regions

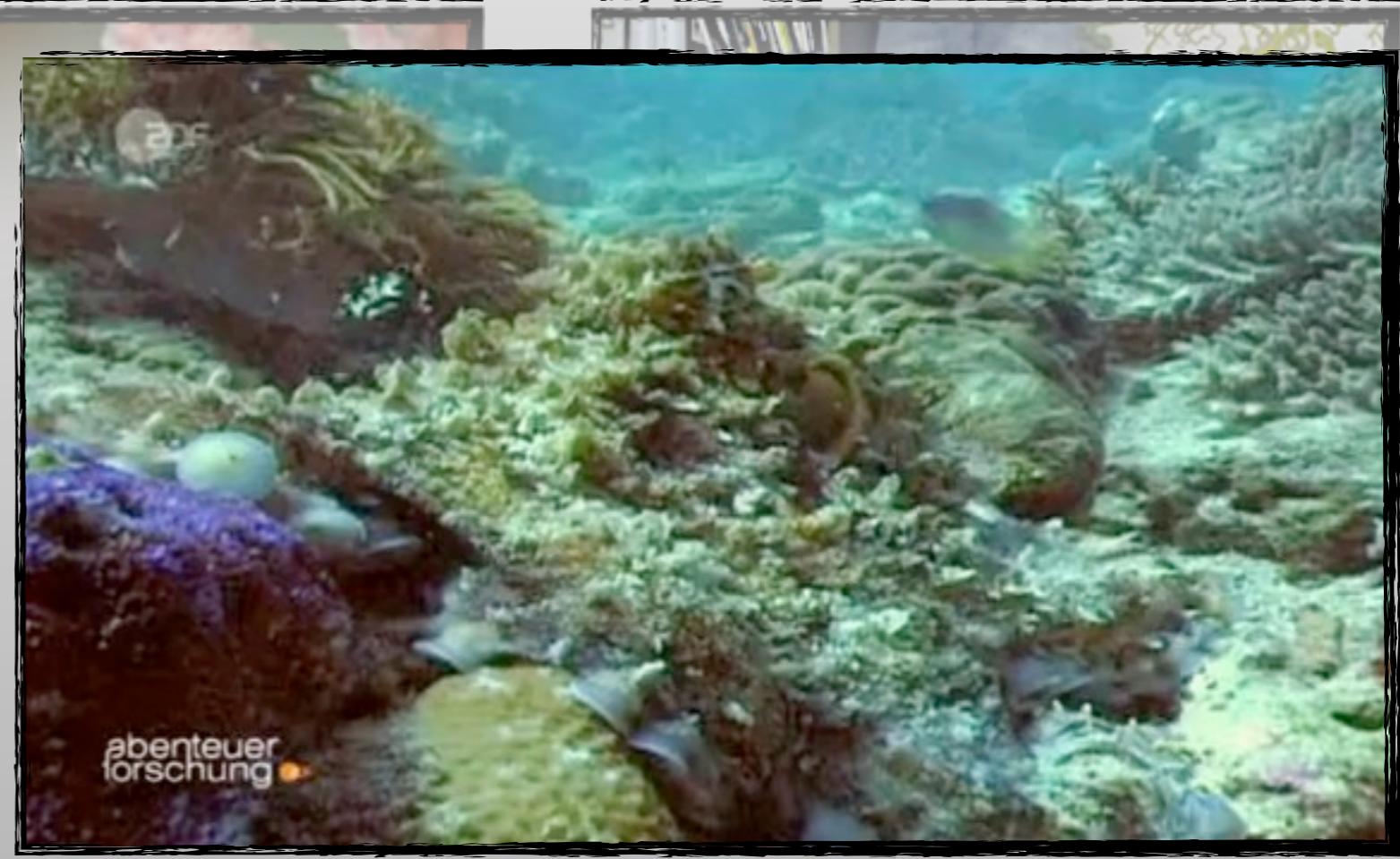
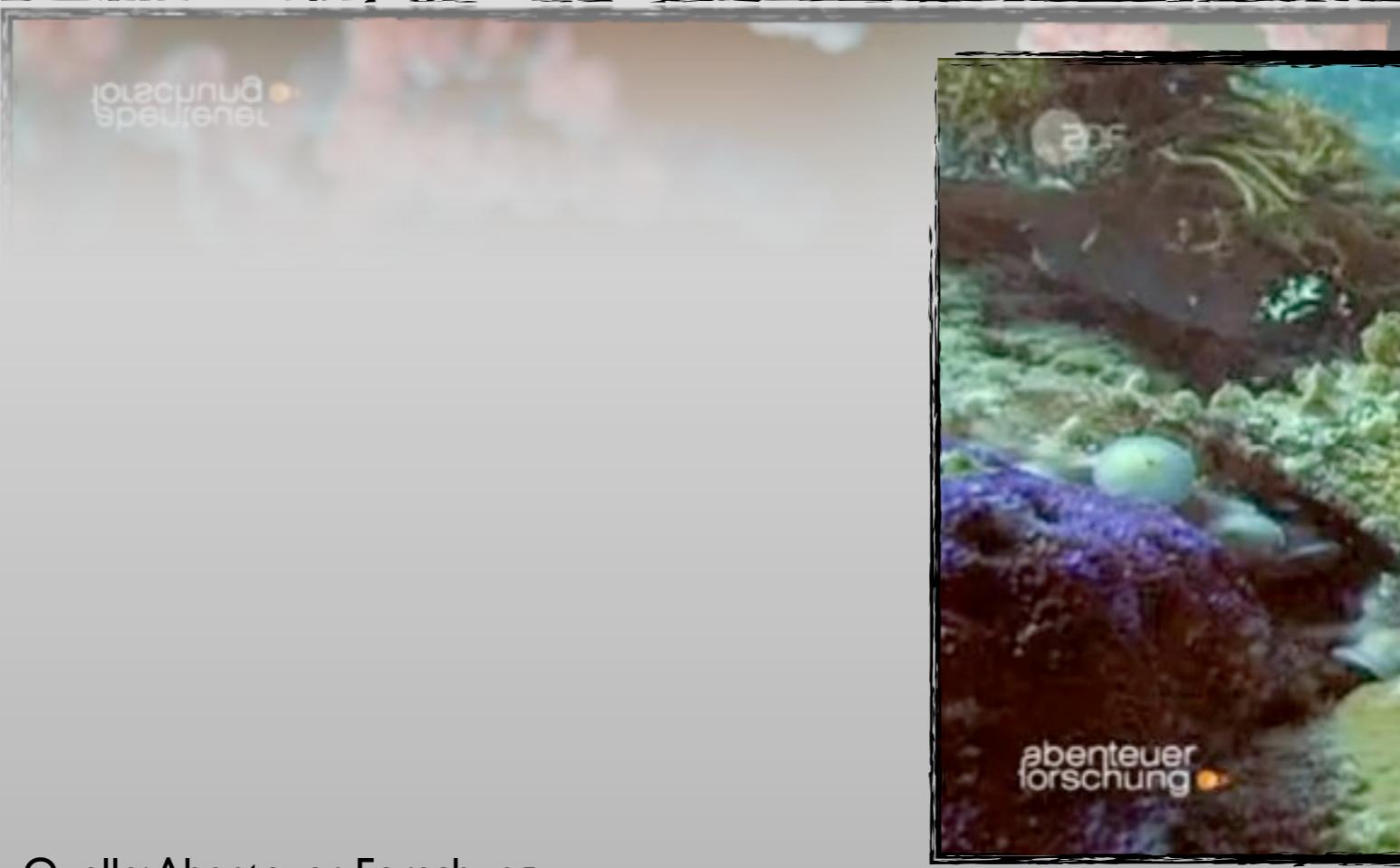


Hier versteckt sich eine Person.
Kannst du sie finden?



Quelle: Abenteuer Forschung

Alles Lug und Trug - Täuschung als Erfolgsrezept - <http://youtu.be/taYkNy-z200>



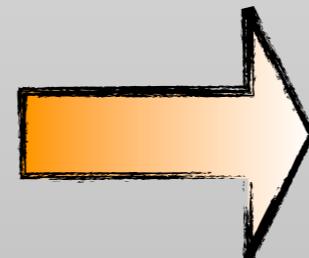
Quelle: Abenteuer Forschung

Alles Lug und Trug - Täuschung als Erfolgsrezept - <http://youtu.be/taYkNy-z200>

- Viele gegenwärtige Ansätze erfordern manuelles Training
- Selbst Säuglinge sind schon in der Lage, Objekte erfolgreich vom Hintergrund zu trennen
- Der Schlüssel: Bewegung!

- Der Schlüssel: Bewegung!

Framework



Optic Flow Tracker

Motion Clustering

Dense Label Interpolation

[1], [2]

[3]

[4]

[1] *Dense point trajectories by GPU-accelerated large displacement optical flow*
N. Sundaram and T. Brox and K. Keutzer (ECCV 2010)

[2] *Large displacement optical flow: descriptor matching in variational motion estimation*
T. Brox and J. Malik (PAMI 2011)

[3] *Object segmentation by long term analysis of point trajectories*
T. Brox and J. Malik (ECCV 2010)

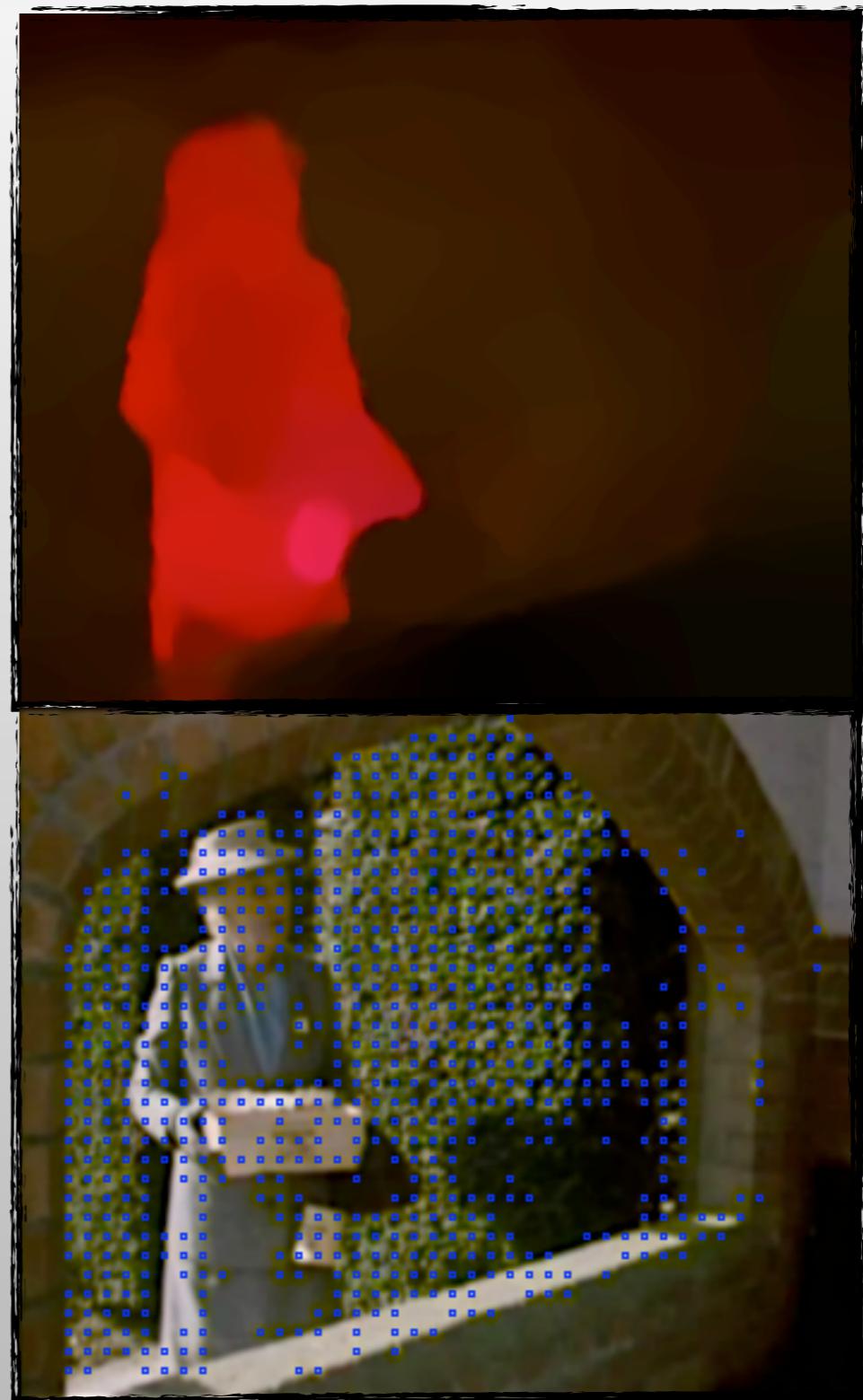
[4] *Object segmentation in video: a hierarchical variational approach for turning point
trajectories into dense regions*
P. Ochs and T. Brox (ICCV 2011)

Optic Flow Tracker

Motion Clustering

Dense Label Interpolation

- berechnet zunächst Large Displacement Optical Flow (LDOF) zwischen zwei Frames
- initialisiert eine Menge von Punkten auf dem ersten Frame
- Punkte auf schwachen Strukturen werden verworfen



WERDEN VERWORFEN

SCHEINBEGEHRUNGEN

Optic Flow Tracker

Motion Clustering

Dense Label Interpolation

- verbleibende Punkte werden entsprechend ihres optischen Flusses zum nachfolgenden Bild verfolgt
- Forward/Backward Check überprüft die Richtigkeit
- Wiederholung der Initialisierung von neuen Punkten in leeren Regionen

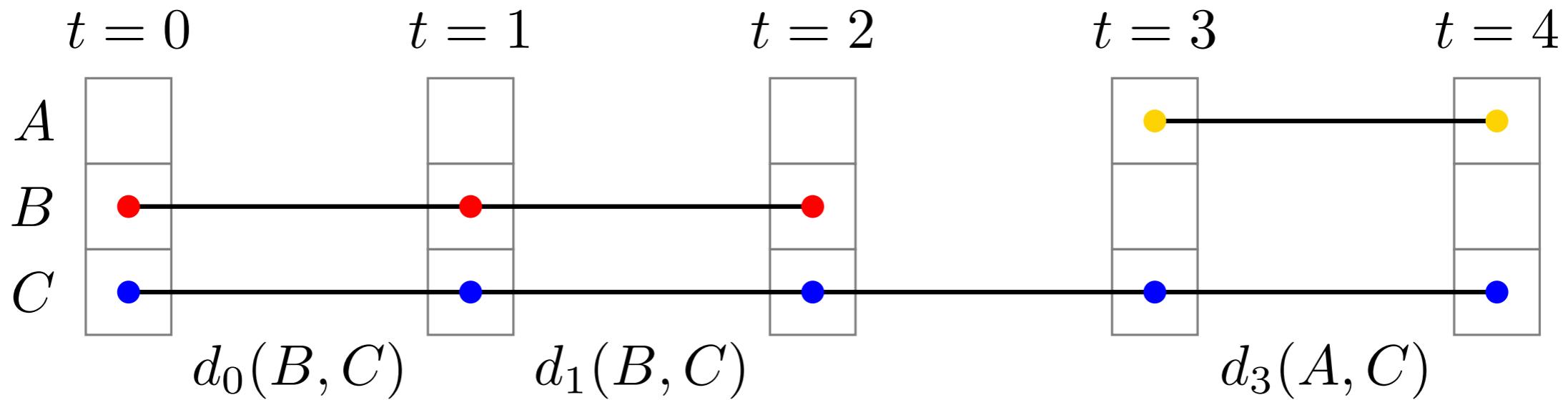


Optic Flow Tracker

Motion Clustering

Dense Label Interpolation

- betrachtet wird das Set von Punkttrajektorien A, B, C, \dots der gesamten Videosequenz
- Definition einer Distanzmetrik zwischen zwei Trajektorien als $d(T_1, T_2) = \max_t d_t(T_1, T_2)$
- berücksichtigt lokalen optischen Fluss, um unterschiedliche Bewegungsgeschwindigkeiten auszugleichen

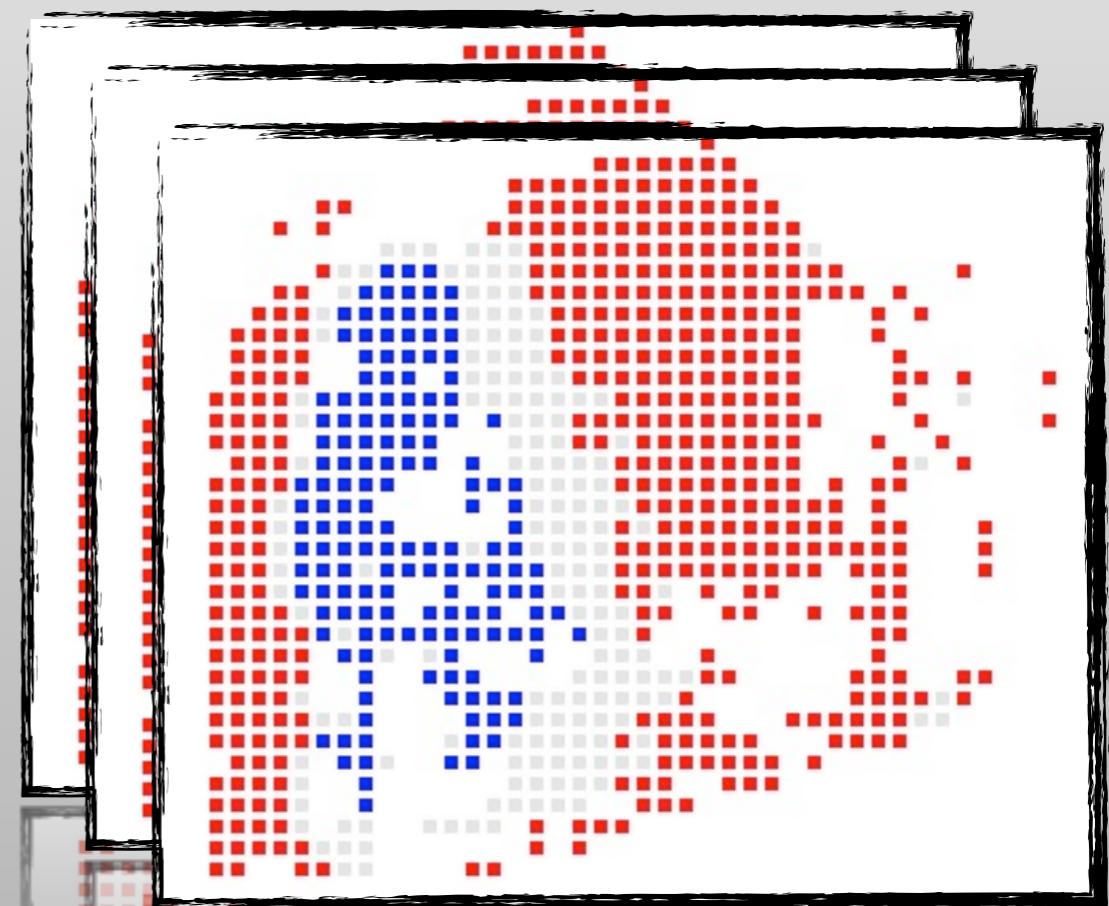
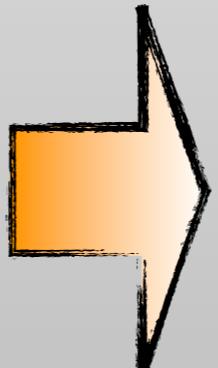


Optic Flow Tracker

Motion Clustering

Dense Label Interpolation

- alle paarweisen Distanzen werden Affinitätsmatrix übertragen
- modifizierte Variante von *Spectral Clustering* bestimmt sowohl Anzahl der Cluster als auch die Zuweisung

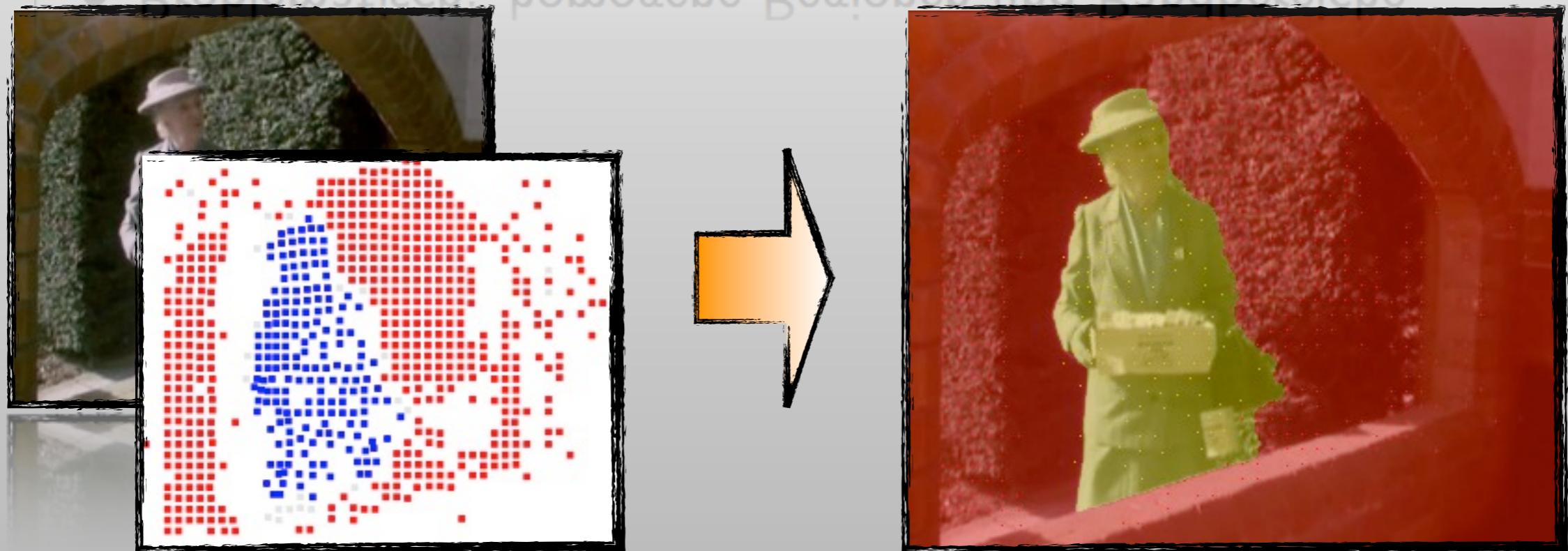


Optic Flow Tracker

Motion Clustering

Dense Label Interpolation

- ausfüllen der nicht klassifizierten Bereiche: $\sim 3\% \rightarrow 100\%$
- Berechnung auf Einzelbildern
- Problematisch: homogene Regionen und Randbereiche



Label Interpolation

- Gesucht: $u := (u_1, \dots, u_n) : \Omega \rightarrow \{0, 1\}^n$
- Energiefunktional minimieren:

$$E(u) = \int_{\Omega} D(u) + \alpha S(u) dx$$

- $D(u)$: Datenterm
- $S(u)$: Glattheitsterm

- $\mathcal{G}(u)$: Glattheitsterm

$$E_{\text{data}}(u) := \int_{\Omega} c \sum_{i=1}^n (u_i - \tilde{u}_i)^2 dx$$

- \tilde{u} : Eingangsdaten
- c : label indikator funktion, d.h. berücksichtigt nur Punkte, die eine Trajektorie besitzen
- lässt die gesuchte Lösung an den groben Eingangsdaten festhalten

$$E_{\text{reg}}(u) := \int_{\Omega} g\psi \left(\sum_{i=1}^n |\nabla u_i|^2 \right) dx$$

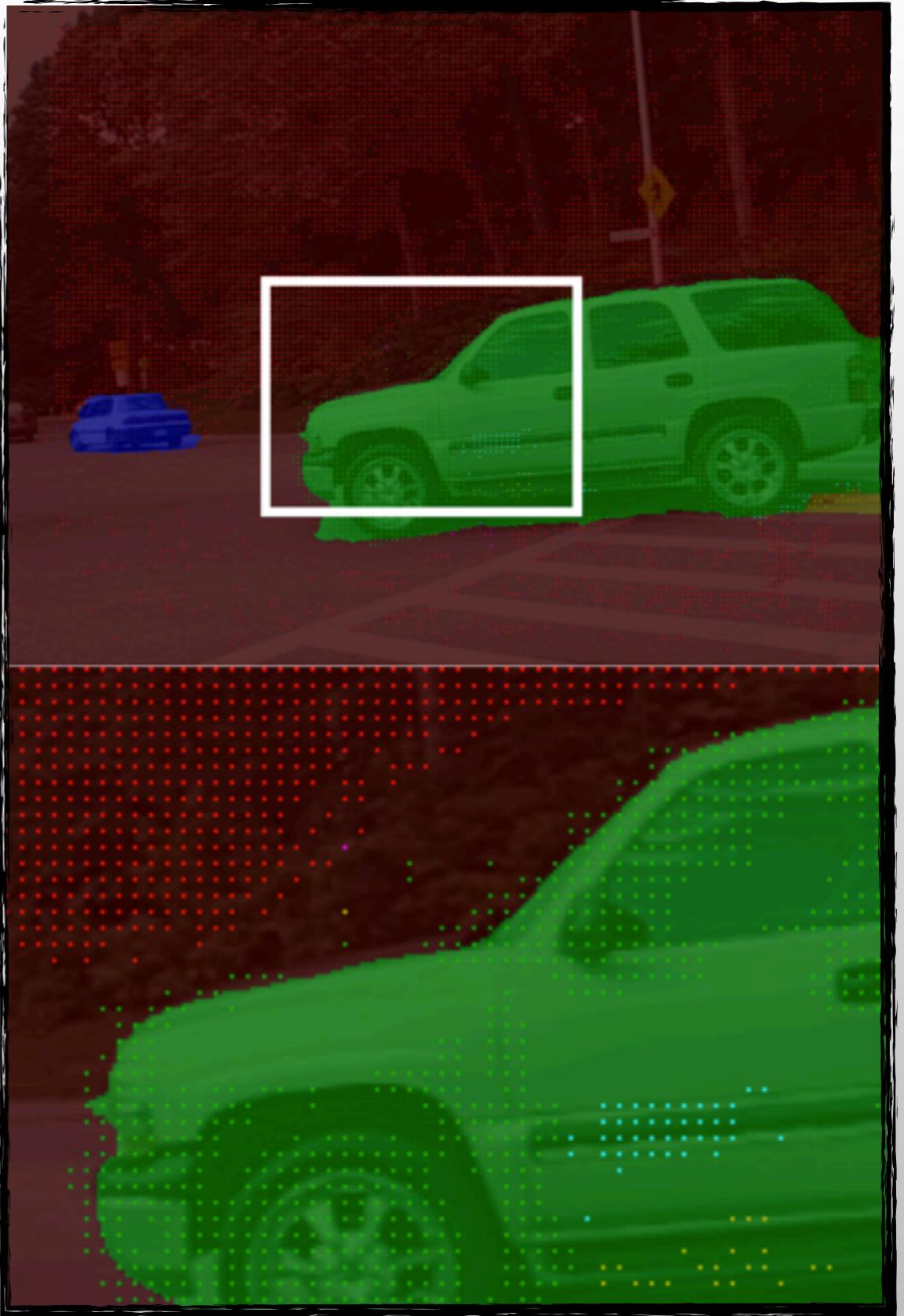
- bevorzugt kompakte Regionen
- $\psi(s^2) := \sqrt{s^2 + \epsilon^2}$
- Ausbreitung bevorzugt in homogene Bereiche

$$g(|\nabla I(x)|^2) := \frac{1}{\sqrt{|\nabla I(x)|^2 + \epsilon^2}}$$

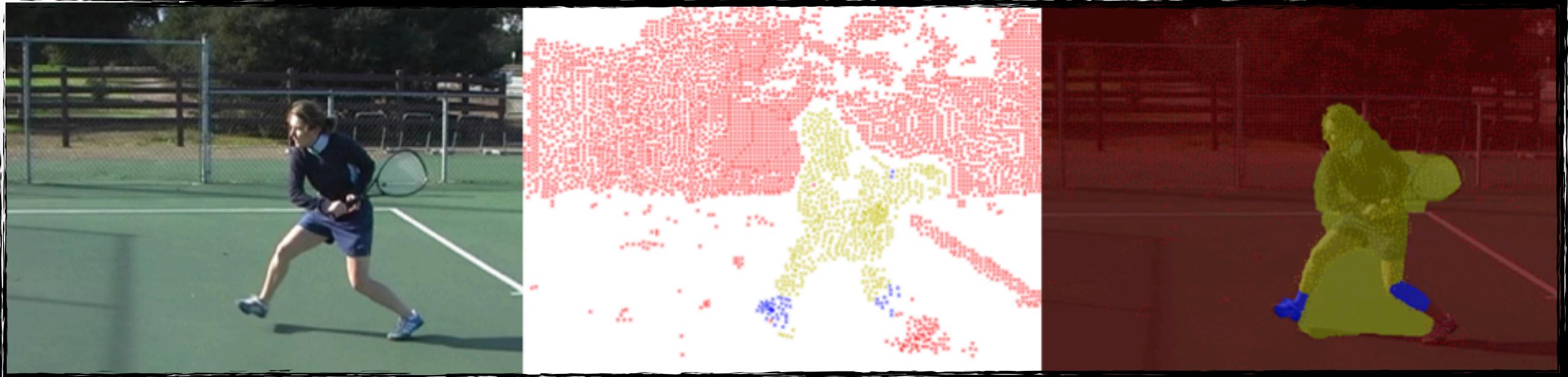
$$\begin{aligned}
E(u) := & \alpha \int_{\Omega} c \sum_{i=1}^n (u_i - \tilde{u}_i)^2 dx \\
& + (1 - \alpha) \int_{\Omega} g \psi \left(\sum_{i=1}^n |\nabla u_i|^2 \right) dx
\end{aligned}$$

- $\alpha \rightarrow 1$: Interpolation, hält an gegebenen Labels fest
- sonst: Approximation, ermöglicht Fehlerkorrektur

- die ursprünglichen Labels dienen als Sourcepunkte
- nichtlineare Diffusion verbreitet die Information in die Nachbarschaft
- je nach Dichte der benachbarten Labels können sich die ursprünglichen Labels am Ende auch ändern



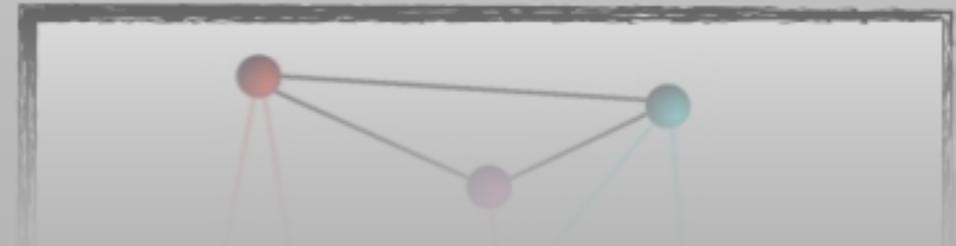
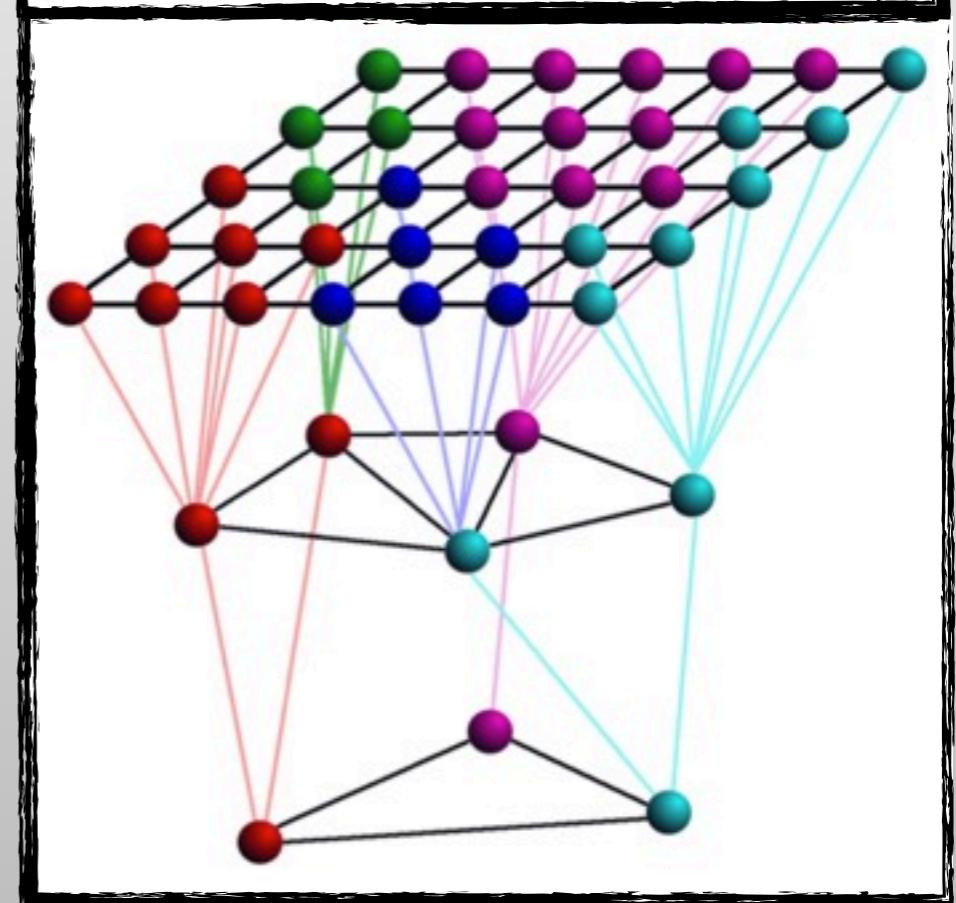
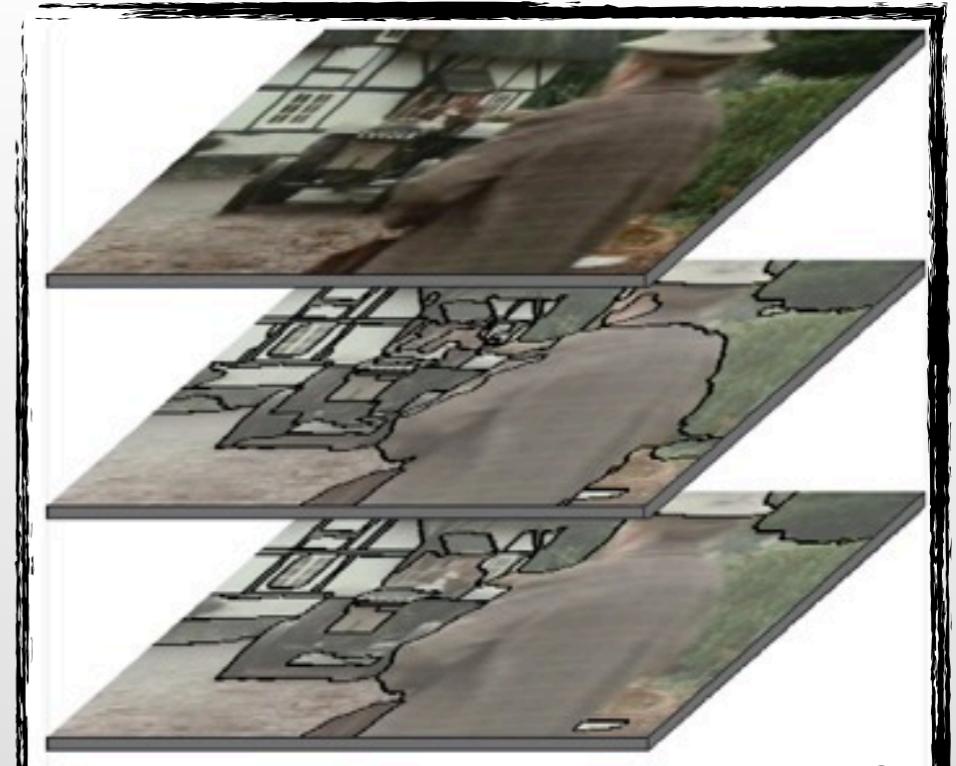
am Ende auch anderen
Gebäuden zugeschlagen



- Problem: geringe Dichte der Sourcepunkte in homogenen Bereichen
- Idee: Verwendung eines hierarchischen Models

• Idee: Verwendung eines hierarchischen Models

- Jede Ebene $k, k = 0, \dots, K$, stellt kontinuierliche Funktion dar
- Feinste Ebene ($k = 0$): Funktionen u^0 und I^0 wie zuvor
- Alle anderen ($k > 0$): stückweise konstante Funktionen u^k und I^k , deren Wert dem Mittelwert des entsprechenden Superpixels entspricht



$$\begin{aligned}
E(\textcolor{red}{u}) := & \alpha \int_{\Omega} c \sum_{i=1}^n (u_i^0 - \tilde{u}_i)^2 dx \\
& + (1 - \alpha) \sum_{k=0}^K \int_{\Omega} g^k \psi \left(\sum_{i=1}^n |\nabla u_i^k|^2 \right) dx \\
& + (1 - \alpha) \sum_{k=1}^K \int_{\Omega} g_l^k \psi \left(\sum_{i=1}^n |u_i^k - u_i^{k-1}|^2 \right) dx
\end{aligned}$$

- Erweiterung der Labelfunktion auf alle Ebenen

$$u := (u_1^0, \dots, u_n^0, u_1^1, \dots, u_n^1, \dots, u_1^K, \dots, u_n^K)$$

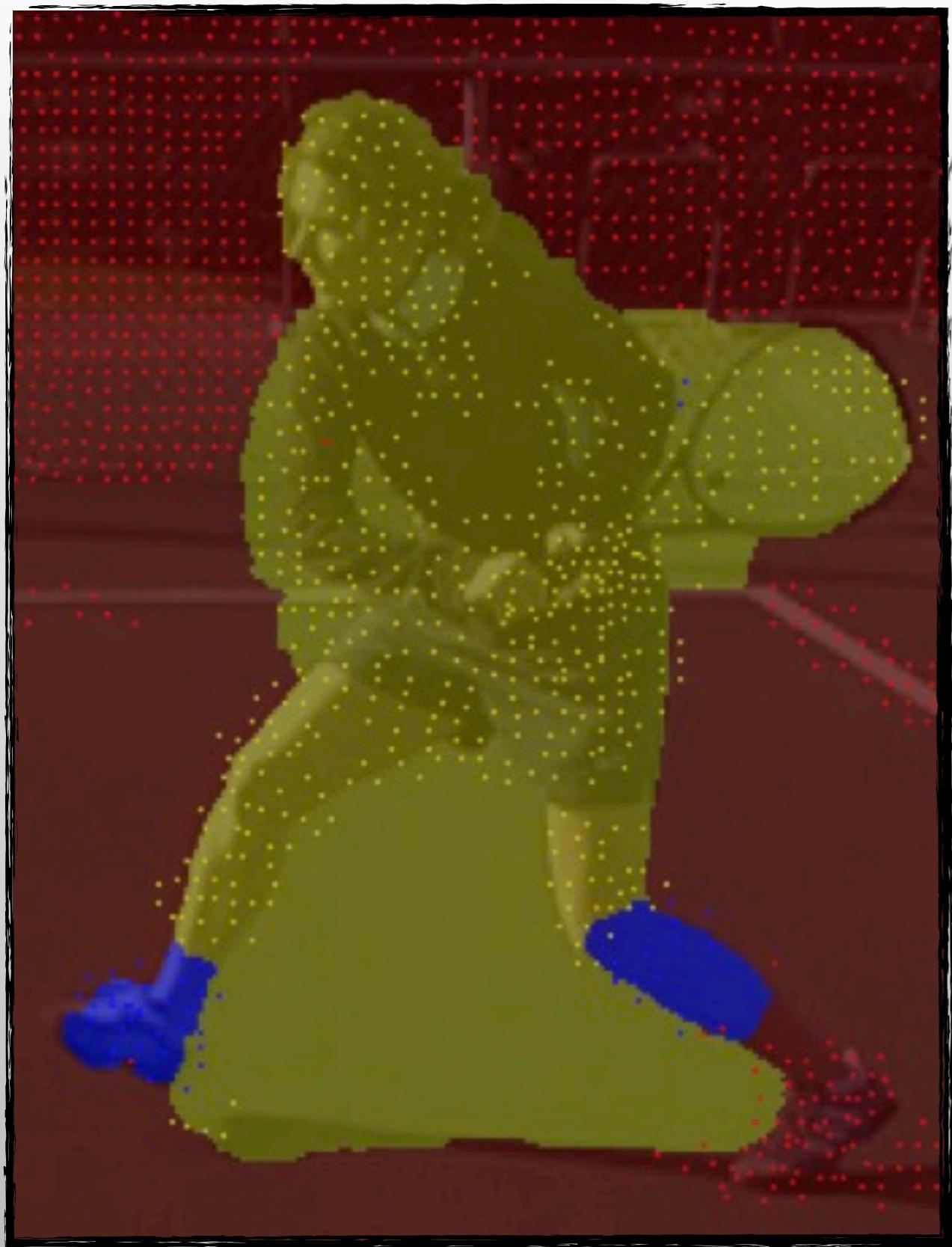
$$\begin{aligned}
E(u) := & \alpha \int_{\Omega} c \sum_{i=1}^n (u_i^0 - \tilde{u}_i)^2 dx \\
& + (1 - \alpha) \sum_{k=0}^K \int_{\Omega} g^k \psi \left(\sum_{i=1}^n |\nabla u_i^k|^2 \right) dx \\
& + (1 - \alpha) \sum_{k=1}^K \int_{\Omega} g_l^k \psi \left(\sum_{i=1}^n |u_i^k - u_i^{k-1}|^2 \right) dx
\end{aligned}$$

- propagiert die Eingangsdaten zu den übergeordneten Ebenen
- gewichtet mit der Ähnlichkeit der Farbwerte

$$g_l^k(x) := \frac{1}{\sqrt{|I^k(x) - I^{k-1}(x)|^2 + \epsilon^2}}$$

$$\begin{aligned}
E(u) := & \alpha \int_{\Omega} c \sum_{i=1}^n (u_i^0 - \tilde{u}_i)^2 dx \\
& + (1 - \alpha) \sum_{k=0}^K \int_{\Omega} g^k \psi \left(\sum_{i=1}^n |\nabla u_i^k|^2 \right) dx \\
& + (1 - \alpha) \sum_{k=1}^K \int_{\Omega} g_l^k \psi \left(\sum_{i=1}^n |u_i^k - u_i^{k-1}|^2 \right) dx
\end{aligned}$$

- Zur Erinnerung: bevorzugt Ausbreitung in homogene Bereiche
- Erweiterung auf alle Ebenen



Der Oktopus





Zusammenfassung

- *Optic Flow Tracker* zur Bestimmung von Punkttrajektorien
- *Motion Clustering* fasst Trajektorien zusammen
- *Dense Label Interpolation*
 - Ausbreitung auf gesamtes Bild unter Berücksichtigung der Farbinformationen
 - Hierarchischer Aufbau, um unterschiedliche Strukturen besser zu erfassen

Strukturen besser zu erfassen

– Hierarchischer Aufbau um unterschiedliche

Anmerkungen

- 64-Bit Linux Binaries und Datensets stehen zur Verfügung:
<http://lmb.informatik.uni-freiburg.de/resources/software.php>
- Hierarchical Image Segmentation: Source Code für Matlab (Linux):
<http://www.eecs.berkeley.edu/Research/Projects/CS/vision/grouping/resources.html>
- System und Testdaten:
 - Intel Core i5-3570K (4x 3.4GHz), 8GB
 - Testbilder: marple2, 50 Bilder x (350x288 Pixel)
- Laufzeiten:
 - LDOF + Tracker + Clustering: 3m 15s für 50 Bilder
 - Hierarchical Image Segmentation: 44m für 50 Bilder
 - Label Interpolation: 8s pro Bild: 6m 30s für 50 Bilder
- Octopus Sequenz mit 100 x (624 x 360): 09:48 Uhr bis 23:09 Uhr (13h)

- Octopus Sequenz mit 100 x (624 x 360): 09:48 Uhr bis 23:09 Uhr (13h)