



Hochschule für Angewandte Wissenschaften Hamburg  
*Hamburg University of Applied Sciences*

# Seminararbeit Anwendungen 2

Tobias Hutzler

Seam Carving - Related Work

Tobias Hutzler  
Seam Carving - Related Work

Seminararbeit im Rahmen der Veranstaltung Anwendungen 2  
im Studiengang Informatik (Master of Science)  
am Department Informatik  
der Fakultät Technik und Informatik  
der Hochschule für Angewandte Wissenschaften Hamburg

Prüfer : Prof. Dr. rer. nat. Kai von Luck

Abgegeben am 28. Februar 2009

# Inhaltsverzeichnis

<b>1</b>	<b>Einleitung</b>	<b>4</b>
1.1	Inhalt der Arbeit . . . . .	4
<b>2</b>	<b>Seamcarving</b>	<b>5</b>
2.1	Problemstellung dieser Arbeit . . . . .	5
2.2	Technische Betrachtung . . . . .	6
2.3	Überblick des Verfahrens . . . . .	6
2.3.1	Schritt 1: Graustufen Konvertierung . . . . .	7
2.3.2	Schritt 2: Kantenerkennung . . . . .	7
2.3.3	Schritt 3: Energiekarte . . . . .	8
2.3.4	Schritt 4: Minimaler Engeriefad . . . . .	9
<b>3</b>	<b>Related Work</b>	<b>11</b>
3.1	Histograms of Oriented Gradients for Human Detection . . . . .	11
3.1.1	Verfahren . . . . .	11
3.1.1.1	Klassifikation von Objekten . . . . .	12
3.1.1.2	Histograms of Oriented Gradients Algorithmus . . . . .	13
3.1.2	Testdaten . . . . .	13
3.1.3	Fazit . . . . .	14
3.2	Gaze Based Interaction for Semiautomatic Photo Cropping . . . . .	15
3.2.1	Verfahren . . . . .	16
3.2.1.1	Photo representation . . . . .	16
3.2.1.2	Content identification . . . . .	16
3.2.1.3	Extracting complete objects . . . . .	17
3.2.1.4	Weitere Schritte des Verfahrens . . . . .	17
3.2.2	Das eye-tracking Verfahren . . . . .	18
3.2.3	Fazit . . . . .	18
<b>4</b>	<b>Zusammenfassung</b>	<b>20</b>
	<b>Abbildungsverzeichnis</b>	<b>21</b>
	<b>Literaturverzeichnis</b>	<b>22</b>

# 1 Einleitung

Diese Ausarbeitung untersucht thematisch angrenzende Gebiete, um die in [Hut09] aufgeführte Vision eines Fotoalbums für mobile Endgeräte durch Seam-Carving zu unterstützen. Hierfür verfolgt der Autor den Einsatz des von Avidan und Shamir auf der SIGGRAPH 2007 vorgestellten Verfahrens [AS07], um den Displaybereich mobiler Endgeräte für die Darstellung von Bildern effektiver als bisher zu nutzen. Trotz der beeindruckenden Leistung von Seam-Carving, weist das Verfahren einige Schwächen auf. Ziel dieser Ausarbeitung ist es Verfahren aufzuzeigen, die Seam-Carving dahingehend unterstützen können, eine Software zu entwickeln, welche die Akzeptanz der Anwender erfährt.

## 1.1 Inhalt der Arbeit

Das Kapitel 2 liefert eine Einführung in Seam-Carving. Hier wird das Vorgehen beschrieben und das Verständnis des Lesers für die in 2.1 definierte Problemstellung geschärft. Das Kapitel 3 „Related Work“ zeigt zwei Verfahren auf, um den Point of Interest (POI) in Bildern zu finden und für weitere Bildbearbeitungen zur Verfügung zu stellen. Abschließend werden die gewonnenen Erkenntnisse in Kapitel 4 festgehalten.

## 2 Seamcarving

Seam-Carving [AS07] ist ein inhaltssensibles Verfahren, um Bilder in Abhängigkeit ihrer Inhalte zu skalieren. Der Ausdruck „*inhaltssensibel*“ ist eine direkte Übersetzung des Fachausdrucks „*content-aware*“. *Content* bezieht sich hierbei auf den Bildinhalt selbst und ist hier also das Bild als technische Information. „Content-aware image resizing“ bezeichnet die Veränderung der Bildgröße nach Analyse der Bilddaten.

### 2.1 Problemstellung dieser Arbeit

Zwischen dieser Art von „*content*“, verstanden als Bilddaten im technischen Sinn, und der Interpretation eines Bildes durch den Betrachter „*user*“ steht, nach Ansicht von Smeulder et al., ein „*semantic gap*“ (vgl. [SWS<sup>+</sup>00]).

*The semantic gap is the lack of coincidence between the information that one can extract from the visual data and the interpretation that the same data have for a user in a given situation.*[SWS<sup>+</sup>00]

Die Betrachtung eines Bildes lässt sich vom Standpunkt des Betrachters aus in *semantic informativeness* und *visual informativeness* aufspalten.

Der *visual informativeness* einer Bildregion entspricht dabei ihrem strukturellen Herausragen aus dem Bild und wird als Salienz bezeichnet. Genau hier, im Zusammenhang zwischen Salienz und Semantik, liegt der Grund für das Funktionieren aber auch das Scheitern von Seam-Carving: Visuelle und inhaltliche Komplexität an einem Ort des Bildes korrelieren oft - aber nicht immer. Dieser Problemstellung widmet sich diese Ausarbeitung. In vielen Fällen kann allein durch die Salienzberechnung herausgefunden werden welche Bereiche für den Bildbetrachter relevant sind und welche nicht. Im Folgenden wird Seam-Carving von seiner algorithmischen Seite betrachtet und weitere Verfahren aufgezeigt, die Anreize schaffen um den „*semantic gap*“ zu überbrücken und damit den Erfolg des Verfahrens steigern.

## 2.2 Technische Betrachtung

Mit Hilfe algorithmisch erzeugter Salienzkarten von Bildern lässt sich beispielsweise vorher-sagen, wo ein Betrachter des Bildes Bereiche mit interessanten Objekten empfindet. Diese Regionen weisen oft eine hohe visuelle Komplexität auf und korrelieren in den meisten Fällen mit den für den Betrachter interessanten Bereichen. Der Seam-Carving-Algorithmus arbeitet grundsätzlich auf den Bildbereichen geringer Energie und operiert somit genau auf den Bereichen des Bildträgers mit geringer Salienz. Dadurch, dass Seam-Carving auf den weniger interessanten Bildbereichen (Bildbereiche mit geringer Energie) operiert, bleibt das Bildobjekt für den Betrachter erhalten. Das Verfahren bleibt inhaltssensibel in semantischer Hinsicht. Avidan und Shamir vereinfachen die Aussage zur Korrelation von Semantik und Salienz zu:

*Intuitively, our goal is to remove unnoticable pixels that blend with their surroundings.* [AS07]

Der folgende Abschnitt widmet sich der algorithmischen Betrachtung des Verfahrens.

## 2.3 Überblick des Verfahrens

Die Hauptkomponente von Seam-Carving ist die Berechnung der Bildenergie eines Bildes. Ziel ist es einen verknüpften Pfad über die Pixel eines Bildes vertikal oder horizontal zu finden, der die geringste Energie aufweist. Der Pfad mit der geringsten Energie entspricht dem Pfad des geringsten Interesses für den Betrachter. Dieser Pfad mit der geringsten Energie kann im Falle einer Verkleinerung ausgeschnitten bzw. im Falle einer Vergrößerung an diesem Pfad ein Seam eingefügt werden.

Folgende Schritte sind für die Berechnung eines Seam notwendig:

1. Das Bild in ein Graustufen Bild konvertieren
2. Auf dem Graustufen Bild ein Kantenerkennungsverfahren ausführen
3. Jeder Wert in der Energiekarte wird mit dem korrespondierenden Kantenwert und dem kleinsten Energiewerten der drei darüber bzw. links liegenden Pixel aufaddiert.
4. Durch „Dynamische Programmierung“<sup>1</sup> wird der minimalste verknüpfte Energiepfad gefunden

---

<sup>1</sup>Wikipedia Eintrag [http://en.wikipedia.org/wiki/Dynamic\\_programming](http://en.wikipedia.org/wiki/Dynamic_programming), referenziert am 27.02.2009



Abbildung 2.1: Originalbild, Quelle: <http://c-a-i-r.wiki.sourceforge.net/Image+Energy>, referenziert am 27.02.2009

### 2.3.1 Schritt 1: Graustufen Konvertierung

Das Originalbild in Abbildung 2.1 wird über folgende Funktion in ein Graustufen Bild 2.2 umgerechnet <sup>2</sup>:

```
1 gray_value[x][y] =  
2 (color[x][y].Red * 0.299) + (color[x][y].Green * 0.587) + (color[x][y].Blue * 0.114)
```

### 2.3.2 Schritt 2: Kantenerkennung

Ist das Bild in ein Graustufen Bild umgerechnet, kann ein Kantenerkennungsalgorithmus eingesetzt werden, um die Bildenergie zu bestimmen. In Sharvin et al. werden anderen Verfahren unter anderem *HoG* und *e1* eingesetzt [AS07]. Ein Kantenerkennungsalgorithmus eignet sich aufgrund seiner einfachen Umsetzung und, im Vergleich zu anderen Methoden, guten Laufzeit. Über die Kantenerkennung lässt sich herausfinden, wo sich Objekte im Bild befinden. Regionen mit vielen Kanten haben einen hohen Energiewert.

Was bedeutet dies nun? Der Kantenwert spiegelt den Übergang von einem Pixel zu seinem benachbarten Pixel wieder. Ein hoher Wert resultiert beispielsweise aus dem Übergang vom tiefblauen Ozean zum linken Astronauten in Abbildung 2.1, wohingegen ein beliebiger Pixel im Ozean zu seinen Nachbarbildpunkt eine sehr geringen Kantenwert aufweist. Kantenerkennungsalgorithmen basieren alle auf dem Faltungsverfahren mit einem entsprechenden Kernel. Die hier vorgestellte Kantenerkennung nutzt eine *Prewitt Kernel* <sup>3</sup>:

<sup>2</sup>Wikipedia Eintrag, <http://en.wikipedia.org/wiki/YUV>, referenziert am 27.02.2009

<sup>3</sup>Wikipedia Eintrag, <http://en.wikipedia.org/wiki/Prewitt>, referenziert am 27.02.2009

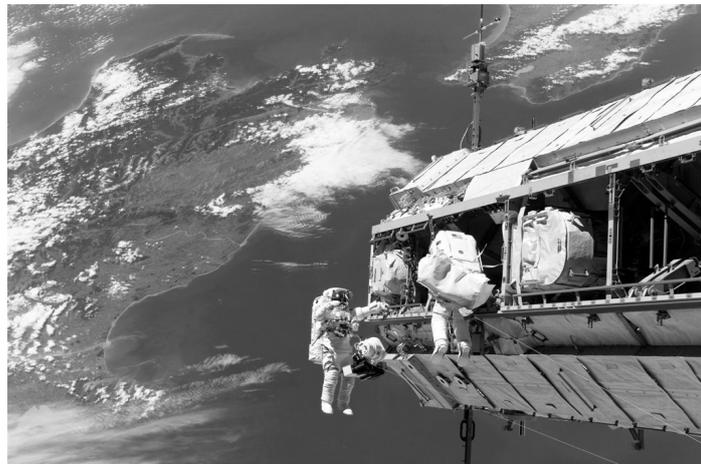


Abbildung 2.2: Das Originalbild als graustufen Bild, Quelle: <http://c-a-i-r.wiki.sourceforge.net/Image+Energy>, referenziert am 27.02.2009

```

1 Prewitt_X = [ [-1, 0, 1 ]
2               [-1, 0, 1 ]
3               [-1, 0, 1 ] ]

5 Prewitt_Y = [ [-1, -1, -1]
6               [ 0,  0,  0]
7               [ 1,  1,  1] ]

```

Die Faltungsfunktion für eine Pixel sieht wie folgt aus:

```

1 edge[x][y] = abs( gray[x+1][y+1] + gray[x+1][y] + gray[x+1][y-1] //X-Anteil des Prewitt
2               -gray[x-1][y-1] - gray[x-1][y] - gray[x-1][y+1] ) +
3               abs( gray[x+1][y+1] + gray[x][y+1] + gray[x-1][y+1] //Y-Anteil des Prewitt
4               -gray[x+1][y-1] - gray[x][y-1] - gray[x-1][y-1] )

```

Da die Intensität des Kantenübergangs ausschlaggebend ist, wird der absolute Betrag der Faltungsfunktion aufaddiert.

### 2.3.3 Schritt 3: Energiekarte

Der entscheidenden Teil von Seam-Carving ist die Erstellung der Energiekarte eines Bildes. Hierzu wird in horizontaler und vertikaler Richtung für jeden Pixel sein Energiewert mit folgender Formel berechnet:

```

1 energy_map[x][y] = minimum( energy_map[x-1][y-1], energy_map[x][y-1], energy_map[x+1][y
2               -1] ) + edge[x][y]

```



Abbildung 2.3: Die Kantekarte des grauskalierten Bildes, Quelle: <http://c-a-i-r.wiki.sourceforge.net/Image+Energy>, referenziert am 27.02.2009

### 2.3.4 Schritt 4: Minimaler Engeriepfad

Um nun den Pfad mit der geringsten Energie zu bestimmen, kommt *dynamische Programmierung*<sup>4</sup> zu Einsatz. Abbildung 2.4 zeigt exemplarisch die Berechnung eines minimalen vertikalen Energiepfades. Es sind nur die fiktiven Energiewerte der Bildpunkte dargestellt.

In der obersten Pixelzeile der Energiekarte wird der Bildpunkt mit der geringsten Energie ausgewählt. Die Berechnung des direkt verknüpften Pfades erfordert nur die Betrachtung der drei anliegenden Bildpunkte in der darunter liegenden Pixelzeile. Gespeichert wird der Bildpunkt mit der geringsten Gewichtung. Dieser Schritt wird so lange wiederholt bis man in der 1. Pixelzeile angekommen sind. Über die eingetragene Kantengewichtung können wir den minimalen Pfad bestimmen. In Abbildung 2.5 ist der minimale Energiepfad in der Energiekarte des Originalbildes rot eingezeichnet.

---

<sup>4</sup>Wikipedia Eintrag [http://en.wikipedia.org/wiki/Dynamic\\_programming](http://en.wikipedia.org/wiki/Dynamic_programming), referenziert am 27.02.2009

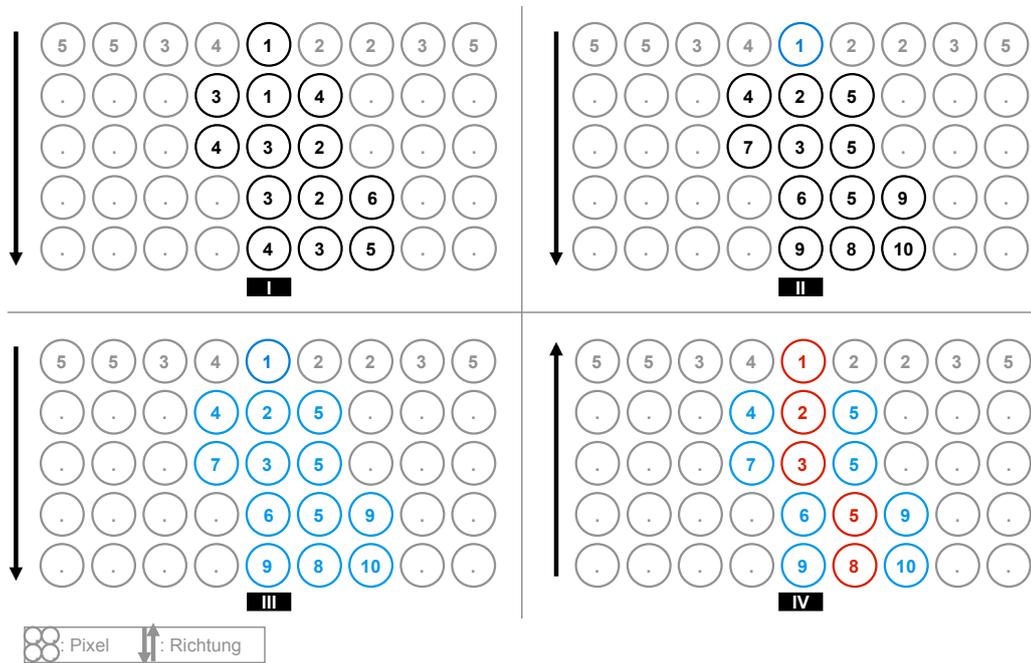


Abbildung 2.4: I zeigt die Energiekarte eines Bildes. In II wird der Bildpunkt (blau) mit dem geringsten Energiewert gewählt. III verdeutlicht die Vorwärtsbestimmung der Nachbarbildpunkte in der nächsten Zeile. Wobei immer der Bildpunkt, dessen Energiewert minimal ist, der Ausgangspunkt für den nächsten Bildpunkt ist. IV zeigt den rückwärts bestimmten minimalen Energiepfad (rot).

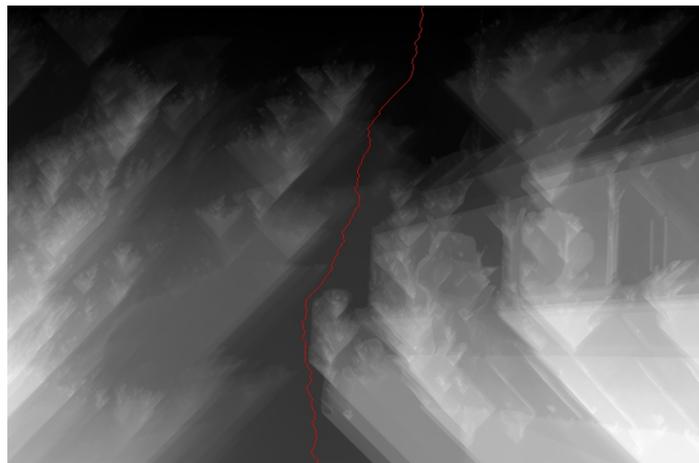


Abbildung 2.5: Energiekarte des Bildes. Minimaler Energiepfad rot eingezeichnet, Quelle: <http://c-a-i-r.wiki.sourceforge.net/Image+Energy>, referenziert am 27.02.2009

## 3 Related Work

Dieses Kapitel zeigt Verfahren auf, die sich mit der Objekterkennung in Bildern beschäftigen. Im Abschnitt 3.1 wird ein Verfahren vorgestellt, das Menschen in Bildern erkennt. Eine weitere Möglichkeit wichtige Regionen in Bildern zu identifizieren, ist mit Hilfe eines Eye-Trackers. Das gaze-based Verfahren mit Blickpunkte und Blickdauer des Bildbetrachters, um auf interessante Bildbereiche zu schließen. Dieses Verfahren wird in zweiten Teil diese Kapitels 3.2 vorgestellt und diskutiert.

### 3.1 Histograms of Oriented Gradients for Human Detection

Personen in Bildern zu erkennen ist nicht nur in sicherheitsrelevanten Anwendungen, wie bei der Einbruchserkennung oder Aktivitätsüberwachung in sicherheitssensiblen Räumen und Orten gefragt, sondern kann auch zu Automatisierungszwecken verwendet werden. Beispielsweise zum Annotieren von Bildern.

Der Autor wird im Folgenden dieses Verfahren vorstellen und auf die Eignung zur Erkennung relevanter Regionen in Bildern diskutieren. Seam-Carving stößt bei Menschen insbesondere bei Gesichtern auf seine Grenzen. Durch das von Dala und Triggs vorgestellte Verfahren [DT05] könnten diese sensitiven Bildbereiche geschützt werden. Geschützt bedeutet hierbei die Energiewerte in den Regions of Interest so zu erhöhen, dass sie durch das Verfahren nicht herangezogen werden.

Im Folgenden wird das Verfahren vorgestellt, auf die Testdaten eingegangen und abschließend die Eignung des Verfahrens für den Schutz interessanter Regionen diskutiert.

#### 3.1.1 Verfahren

Das Verfahren gliedert sich in zwei Phasen: das Trainieren der Objektklassifizierung und das spätere Finden von Objekten. Zur Klassifizierung wird Support Vector Maschine (SVM) <sup>1</sup> eingesetzt. SVM liefert als Ergebnis „Objekt gefunden“ oder „Objekt nicht gefunden“.

---

<sup>1</sup>Wikipedia Eintrag [http://en.wikipedia.org/wiki/Support\\_vector\\_machine](http://en.wikipedia.org/wiki/Support_vector_machine), referenziert am 27.02.2009

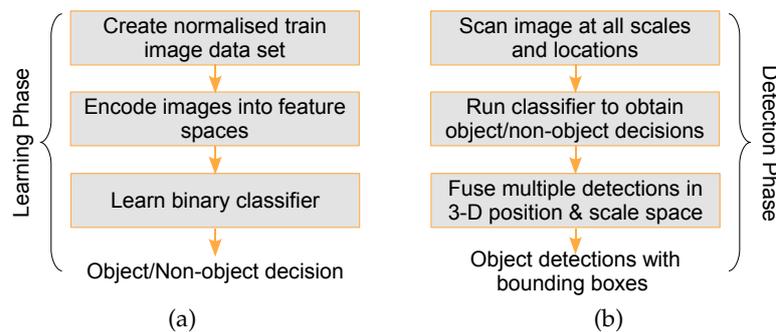


Abbildung 3.1: Überblick Objekterkennungsarchitektur [Dal06]

### 3.1.1.1 Klassifikation von Objekten

**Trainingsphase** In der Trainingsphase wird der SVM Klassifizierer mit positiven und negativen Testdaten trainiert.

**Positive Testdaten** Positive Trainingsbilder sind Bilder in einer bestimmten Größe in deren Zentrum sich das zu klassifizierende Objekt befindet. In diesem Fall eine Person

**Negative Testdaten** Negative Testbilder sind Bilder derselben Größe in denen sich Objekte befinden, welche nicht der Klasse entsprechen, z. B. Bäume, Türen, Fahrzeuge oder Tiere.

**Objekterkennung** Die Objekterkennung lässt sich wiederum in zwei Hauptarbeitschritte unterteilen.

- Über das Eingabebild werden mehrere Erkennungsfenster gelegt in denen dann der Feature-Vektor berechnet wird.
- Der SVM Klassifizierer wird dann mit den entsprechenden Feature-Vektoren gefüttert. Der Vergleich liefert entweder einen positiven oder negativen Treffer. Da sich bestimmte Erkennungsfenster überlagern können werden diese zusammengeführt.

Der Feature-Vektor wird mit der Histograms of Oriented Gradients (*HoG*) [DT05] Methode berechnet, welche im nächsten Abschnitt beschrieben wird.

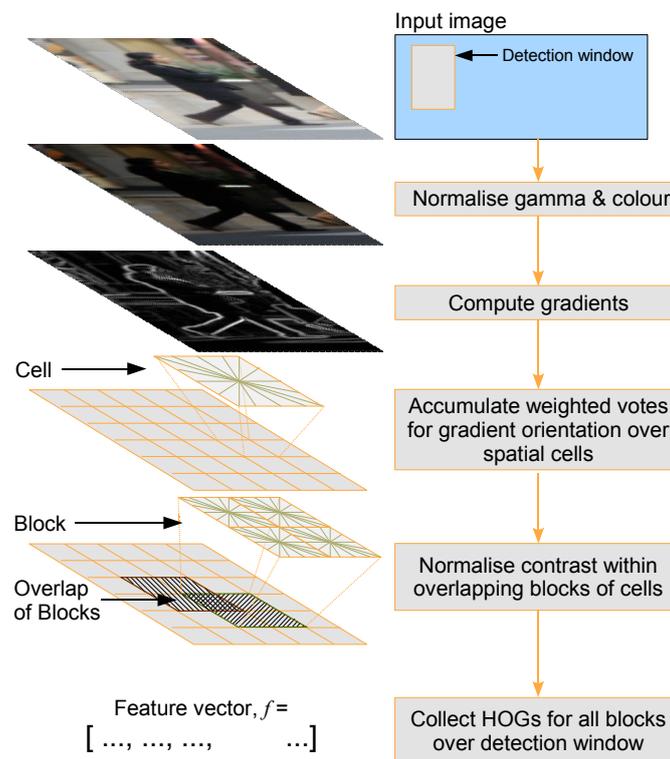


Abbildung 3.2: Zusammenfassung des HOG Algorithmus [Dal06]

### 3.1.1.2 Histograms of Oriented Gradients Algorithmus

HOG basiert auf der Prämisse, dass Objekte in Bildern aufgrund ihrer Intensitätsverteilung der Gradienten oder Kantenrichtung beschrieben werden können. Hierzu wird das Bild in räumliche Bereiche „cells“ eingeteilt. Für jede Zelle wird ein Gradientenhistogramm berechnet. Die Kombination der Histogramme repräsentiert die Objektbeschreibung.

Abbildung 3.2 zeigt einen Überblick über den HOG Algorithmus.

### 3.1.2 Testdaten

Als Test- und Trainingsdaten dienen zwei Bilddatenbanken. Die bekannte Bilddatenbank des MITs<sup>2</sup> zeigt Fußgänger entweder von vorn oder hinten in aufrechter Position in alltäglichen Situationen siehe Abbildung 3.3. Sie beinhaltet 924 Bilder von Fußgängern. Da die Evaluation auf Basis der MIT Datenbank sehr gute Ergebnisse geliefert hatte, entschieden sich

<sup>2</sup>MIT CBCL Pedestrian Database <http://cbcl.mit.edu/software-datasets/PedestrianData.html>, referenziert am 27.02.2009



Abbildung 3.3: Beispielbilder der MIT Pedestrian Data set [Dal06]



Abbildung 3.4: Beispielbilder der INRIA Static Person Data Set [Dal06]

Dalal und Triggs dafür eine weitaus anspruchsvollere Bilddatenbank<sup>3</sup> zu erstellen. Diese beinhaltet weit aus komplexere Szenarien siehe Abbildung 3.4. Es wurden über acht Monate hinweg 497 Bildern von Menschen in den verschiedensten Posen, Wetterbedingungen und Szenarien (Menschenmassen) gesammelt. Diese Bilder weisen gegenüber der MIT Bilderdatenbank eine sehr hohe Varianz auf. Auch hier lieferte das vorgestellte Verfahren sehr gute Ergebnisse im Vergleich zu bisherigen Personenerkennungsverfahren.

### 3.1.3 Fazit

Das Verfahren liefert besonders gute Ergebnisse in der Erkennung von Person in Bildern und ist darüber hinaus nicht auf diese Klasse beschränkt. Das Verfahren wurde experimentell auf folgende Objektklassen der PASCAL Visual Object Challenge [EVGW<sup>+</sup>] angewandt: Fahrräder, Busse, Autos, Motorräder, Katzen, Kühe, Hunde, Schafe, Pferde und Menschen. Die Parameter für die Feature-Set-Berechnung mussten dabei nur geringfügig angepasst werden. Allerdings versagte das Verfahren bei der Erkennung von Katzen (siehe Abbildung 3.5). Gründe hierfür sind zum einen, dass die Objektklassifizierung eine sehr gute Trainingsbilddatenbank voraussetzt, was oft mit sehr hohem Aufwand verbunden ist. Ein weiteres

<sup>3</sup>LEAR <http://lear.inrialpes.fr/data>, referenziert am 27.02.2009

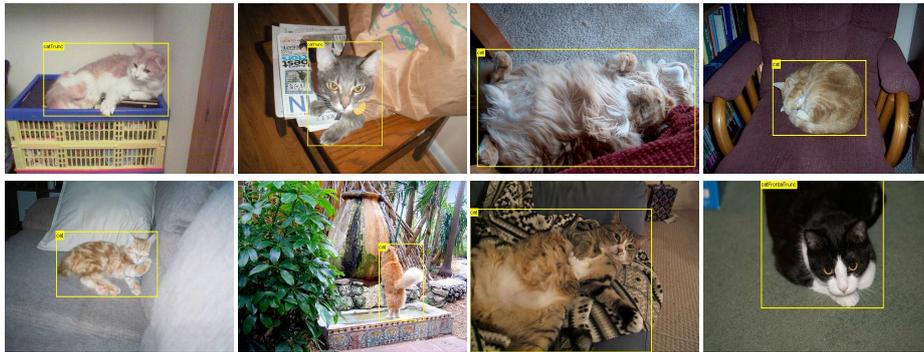


Abbildung 3.5: Trainingsbilder der Klasse für Katzen [Dal06]

Problem, welches durch die Klassifizierung von Katzen deutlich wird, ist die extrem variierende geometrische Form von Katzen (siehe Abbildung 3.5). Vergleicht man diese Bilder mit Motorrädern oder aufrecht stehenden Personen wird dies noch deutlicher.

Der rechenintensivste Teil des Verfahrens ist die Berechnung des HOG Descriptors. Durch einen effizienten Berechnungsalgorithmus wie beispielsweise ADSBoost [Sch] kann die Berechnung nahezu in Echtzeit erfolgen.

Anzumerken ist auch noch die hohe Anzahl an Optimierungsparametern die das Verfahren mit sich bringt. Allerdings konnte festgestellt werden, dass für bestimmte Objektklassen wie Fahrzeuge (z.B. Auto, Bus, Motorrad und Fahrrad) die Parameter sich nur geringfügig unterscheiden (vgl. [Dal06]).

Da dieses Verfahren eines der beiden favorisierten Verfahren von Avidan und Shamir für Seam-Carving ist, eignet sich der Einsatz von *HoG* besonders [AS07]. Die Unterstützung verschiedener Objektklassen des Verfahrens ist ein weitere Vorteil. Zu klären bleibt allerdings noch, in wie weit es sich als mobile Anwendung umsetzen lässt.

## 3.2 Gaze Based Interaction for Semiautomatic Photo Cropping

Santalla et al. [SAD<sup>+</sup>06] stellen ein Verfahren vor, um den wesentlichen Bildausschnitt eines Bildes unter Verwendung eines Eye-Trackers festzustellen. Der Schwerpunkt der Interaktion des Bildbetrachters liegt auf dessen Betrachtung des Bildes. Anhand der Blickpunkte und Blickdauer des Betrachters wird der Point of Interest (POI) bestimmt und über ein algorithmisches Verfahren welches im Weiteren vorgestellt wird, der Ausschnitt in Größe und Position berechnet.

Das Verfahren ermöglicht Anwendungen, welche semi-automatische Schnappschüsse er-

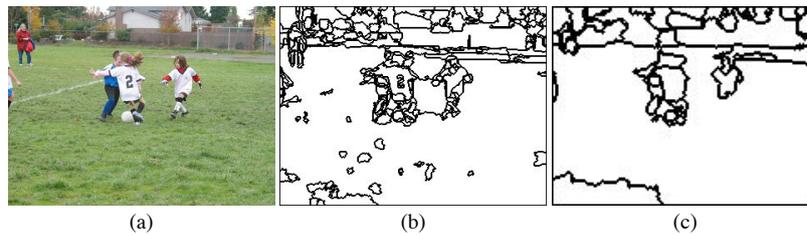


Abbildung 3.6: a) Original Bild b)  $S_{fine}$  c)  $S_{coarse}$

stellen, anpassbare Dokumente zu erzeugen, oder das Thumbnailing von Bildern zu vereinfachen.

### 3.2.1 Verfahren

Das Verfahren ist in mehreren Schritten beschrieben, wobei sich die letzteren ausschließlich auf die Optimierung eines Ausschnittes beziehen. Diese sind für den Autor als Unterstützung für Seam-Carving uninteressant und werden nicht weiter ausgeführt.

#### 3.2.1.1 Photo representation

Um Bildausschnitte zu wählen, die das Wesentliche eines Fotos repräsentieren, müssen die Bildregionen gefunden werden, die möglichst intensiv betrachtet wurden. Das Foto wird dazu fein und grob segmentiert ( $S_{fine}$  und  $S_{coarse}$ ). Zur Segmentierung verwenden Santalla et. al. das Werkzeug EDISON [Geo02]. Die Abbildung 3.6 zeigt das Ergebnis der Fein- und Grobsegmentierung.

#### 3.2.1.2 Content identification

Um den POI eines Bildes zu bestimmen werden Segmente ausgewählt, welche sehr interessant für den Betrachter sind. Die Blickdauer des Betrachters in eine bestimmte Region würde zu ungenaue Ergebnisse liefern. Dies liegt zum einen an den Segmentierungsverfahren, die oft nicht alle Segmentgrenzen eines Bildes genau bestimmen können, zum anderen liegt es aber auch an den Ungenauigkeit von Eye-Trackern, die technisch bedingt einen gewissen Fehler des Blickpunktes und in der Blickdauer aufweisen.

Um diese technischen Missstände zu umgehen wird eine Feinabstimmung zwischen den anliegenden Segmenten und dem Blickpunkt erstellt. Die Blickpunkte und Blickdauer in eine bestimmte Region (Segment) steigern den Interessengrad dieser Region. Interessant ist an dieser Stelle, das Santalla et.al. durch eine Mittlung der Werte erreichen, dass wirklich nur

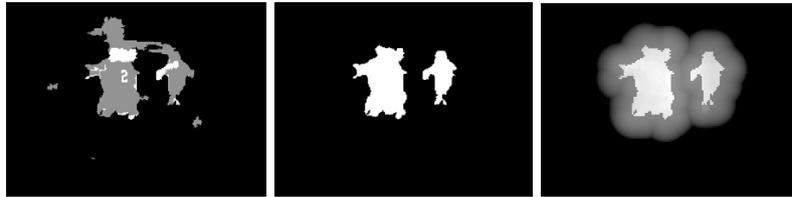


Abbildung 3.7: Dargestellt ist das lazy snapping Verfahren. Links zu sehen sind die Vordergrund-Bereiche (weiß) Hintergrundbereiche (schwarz) nicht zuzuordnende Bereiche (grau). In der Mitte sind nur die Vorder- und Hintergrundregionen dargestellt. Rechts ist die übersegmentierte Inhaltskarte zu sehen. [SAD<sup>+</sup>06]

interessante Objekte in einem Bild erfasst werden. Bei der Betrachtung von Bildern neigt man oft in die Ferne, Bildtiefe zu sehen. Konkret den Himmel oder die Landschaft neben den interessanten Objekten länger zu betrachten. Durch diese Mittlung der Werte wird eine Verschiebung des POI verhindert.

### 3.2.1.3 Extracting complete objects

Durch die Segmentierung setzen sich ganze Objekte oft aus kleineren Segmenten zusammen. Dies ist in Abbildung 3.6 gut zu erkennen. Das Fußball spielende Kind besteht sowohl in der Feinsegmentierung als auch in der Grobsegmentierung aus mehreren Segmenten. Um dennoch ganze Objekte aus mehreren einzelnen Segmenten zu erkennen verwenden Santalla et. al. das „lazy snapping“ Verfahren [LSTS04]. Das Verfahren basiert auf einer Übersegmentierung und teilt das Bild in Vorder- und Hintergrundbereiche ein. Abbildung 3.7 zeigt das Ergebnis der Fußball spielenden Kinder.

### 3.2.1.4 Weitere Schritte des Verfahrens

Die weiteren Schritte des Verfahrens behandeln die Faktoren für einen optimalen Bildausschnitt. Diese beinhalten die Optimierung der Ausschnittsgröße und der Zentrierung des POI im Ausschnitt, sorgen für die Vermeidung, dass Objekte im Hintergrund durchgeschnitten werden.

Die für Seam-Carving wesentlichen Schritte wurden hier aufgezeigt, so dass eine genauere Betrachtung der Ausschnittsoptimierung nicht weiter notwendig ist.

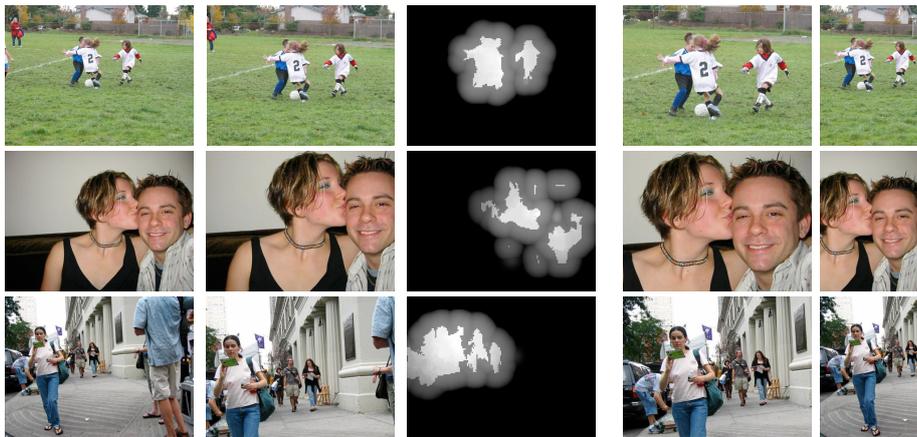


Abbildung 3.8: I Spalte: Originalbild II Spalte: Vollautomatischer Ausschnitt nach [SLBJ03] III Spalte: gaze-based content map IV, V Spalte: gaze-based Ausschnitt horizontale Ausrichtung und vertikale Ausrichtung des Bildausschnittes

### 3.2.2 Das eye-tracking Verfahren

Von Autoren wurden 50 Bilder ausgesucht, die von einem Bildausschnitt profitieren können. Die Bilder zeigten Menschen, Stillleben, Straßenszenen usw. Die Testpersonen wurden davon unterrichtet, dass ihre Augenbewegung zum Ausschneiden des Bildes verwendet wird, allerdings wurde ihnen nicht gesagt, dass sie auf den wesentlichen Bildinhalt blicken sollen. Über Details des Verfahrens wurden die Testpersonen allerdings nicht unterrichtet. Die Augenbewegung wurde für jedes Bild für 10 sec mit einer Tobii x50 registriert. Ergebnisse dieses Verfahrens zeigt folgende Abbildung 3.8.

Der Vergleich mit dem von [SLBJ03] eingesetzten Verfahren zur voll automatischen Thumbnail-Erstellung zeigt die Vorteile des gaze-based Verfahrens auf.

### 3.2.3 Fazit

Das Verfahren zeichnet sich besonders durch eine minimale Interaktion mit dem Benutzer aus, um relevante Bildbereiche eines Fotos zu bestimmen. Der „semantic gap“ des Seam-Carving Verfahrens, welcher in der Problemstellung dieser Arbeit verdeutlicht wurde könnte damit auf ein Minimum verkleinert werden. Interessant ist auch die Mittlung der Blickpunkte und Dauer in Bezug auf Bildhintergründe. Mobil wird es nicht möglich sein eye-tracking einzusetzen. Allerdings sind die Ansätze weiter zu verfolgen. Besonders interessant könnte die Übersegmentierung des lazy snapping Verfahrens in Verbindung mit einer Benutzerinterak-

tion. Hierbei könnten interessante Bereiche durch Drücken auf das Bild an Stelle des Eye Trackings markiert und somit geschützt werden.

## 4 Zusammenfassung

Die Seminararbeit hat eingehend Seam-Carving vorgestellt. In der Problemstellung wurden die Schwachstellen des Verfahrens verdeutlicht. Die Energiekarte des Bildes wird aufgrund der Reliefdaten des Bildes bestimmt. In vielen Fällen besteht eine Korrelation zwischen interessanten Bildbereichen und den Salienzen. Oft aber auch nicht.

In Kapitel 3 wurden Ansätze vorgestellt, die sich im Kern mit der Identifizierung wichtiger Bildbereiche eines Fotos beschäftigen. Das 1. Verfahren hatten den Schwerpunkt auf der Objekterkennung, konkret der Erkennung von Menschen in Bildern, konnte aber auch aufzeigen, dass es für eine Vielzahl anderer Objektklassen funktioniert. Schwachstelle dieses Verfahrens sind Objekte die auf Grund ihrer geometrischen Form in verschiedenen Bildsituationen sich nur sehr schwer klassifizieren lassen, um nur als Beispiel die Klasse Katzen aufzuführen. Das gaze-based Verfahren überbrückt aufgrund der Blickpunkte und Blickdauer des Betrachters den semantic-gap zwischen der technischen und semantischen Repräsentation von Bildern sehr gut. Allerdings ist dazu ein eye-tracker, der die Blickpunkte des Bildbetrachters aufzeichnet, erforderlich. Die Segmentierung, welche in diesem Verfahren eingesetzt wird, könnte sich in Kombination mit einer punktaktigen Markierung der wichtigen Bildbereiche durch den Anwender gut für eine Bereichsschätzung eignen.

Die seit 2005 jährlich stattfindende PASCAL Visual Object Challenge [EVGW<sup>+</sup>] zeigt die wissenschaftliche Aktualität und Nachfrage nach Objekterkennung in Bildern auf. War die Herausforderung des Wettbewerbs 2006 noch auf 10 Objektklassen beschränkt, so wurde der Wettbewerb bereits 2008 auf 20 Objektklassen ausgeweitet. Die Ansätze und Ideen der Wettbewerbsteilnehmer können auf ihre Eignung für die Unterstützung für Seam-Carving untersucht werden.

# Abbildungsverzeichnis

2.1	Originalbild, Quelle: <a href="http://c-a-i-r.wiki.sourceforge.net/Image+Energy">http://c-a-i-r.wiki.sourceforge.net/Image+Energy</a> , referenziert am 27.02.2009 . . . . .	7
2.2	Das Originalbild als graustufen Bild, Quelle: <a href="http://c-a-i-r.wiki.sourceforge.net/Image+Energy">http://c-a-i-r.wiki.sourceforge.net/Image+Energy</a> , referenziert am 27.02.2009 . . . . .	8
2.3	Die Kantekarte des grauskalierten Bildes, Quelle: <a href="http://c-a-i-r.wiki.sourceforge.net/Image+Energy">http://c-a-i-r.wiki.sourceforge.net/Image+Energy</a> , referenziert am 27.02.2009 . . . . .	9
2.4	I zeigt die Energiekarte eines Bildes. In II wird der Bildpunkt (blau) mit dem geringsten Energiewert gewählt. III verdeutlicht die Vorwärtsbestimmung der Nachbarbildpunkte in der nächsten Zeile. Wobei immer der Bildpunkt, dessen Energiewert minimal ist, der Ausgangspunkt für den nächsten Bildpunkt ist. IV zeigt den rückwärts bestimmten minimalen Energiepfad (rot). . . . .	10
2.5	Energiekarte des Bildes. Minimaler Energiepfad rot eingezeichnet, Quelle: <a href="http://c-a-i-r.wiki.sourceforge.net/Image+Energy">http://c-a-i-r.wiki.sourceforge.net/Image+Energy</a> , referenziert am 27.02.2009 . . . . .	10
3.1	Überblick Objekterkennungsarchitektur [ <a href="#">Dal06</a> ] . . . . .	12
3.2	Zusammenfassung des HOG Algorithmus [ <a href="#">Dal06</a> ] . . . . .	13
3.3	Beispielbilder der MIT Pedestrian Data set [ <a href="#">Dal06</a> ] . . . . .	14
3.4	Beispielbilder der INRIA Static Person Data Set [ <a href="#">Dal06</a> ] . . . . .	14
3.5	Trainingsbilder der Klasse für Katzen [ <a href="#">Dal06</a> ] . . . . .	15
3.6	a) Original Bild b) $S_{fine}$ c) $S_{coarse}$ . . . . .	16
3.7	Dargestellt ist das lazy snapping Verfahren. Links zu sehen sind die Vordergrund-Bereiche (weis) Hintergrundbereiche (schwarz) nicht zuzuordnende Bereiche (grau). In der Mitte sind nur die Vorder- und Hintergrundregionen dargestellt. Rechts ist die übersegmentierte Inhaltskarte zu sehen. [ <a href="#">SAD<sup>+</sup>06</a> ] . . . . .	17
3.8	I Spalte: Originalbild II Spalte: Vollautomatischer Ausschnitt nach [ <a href="#">SLBJ03</a> ] III Spalte: gaze-based content map IV, V Spalte: gaze-based Ausschnitt horizontale Ausrichtung und vertikale Ausrichtung des Bildausschnittes . . . . .	18

# Literaturverzeichnis

- [AS07] AVIDAN, SHAI und ARIEL SHAMIR: *Seam carving for content-aware image resizing*. ACM Trans. Graph., 26(3):10, 2007.
- [Dal06] DALAL, NAVNEET: *Finding people in images and videos*. Doktorarbeit, Institut National Polytechnique de Grenoble, July 2006.
- [DT05] DALAL, NAVNEET und BILL TRIGGS: *Histograms of Oriented Gradients for Human Detection*. In: *CVPR '05: Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05) - Volume 1*, Seiten 886–893, Washington, DC, USA, 2005. IEEE Computer Society.
- [EVGW<sup>+</sup>] EVERINGHAM, M., L. VAN GOOL, C. K. I. WILLIAMS, J. WINN und A. ZISSERMAN: *The PASCAL Visual Object Classes Challenge 2007 (VOC2007) Results*. <http://www.pascal-network.org/challenges/VOC/voc2007/workshop/index.html>.
- [Geo02] GEORGESCU, CHRISTOPHER M.: *Synergism in low level vision*. In: *In International Conference on Pattern Recognition*, Seiten 150–155, 2002.
- [Hut09] HUTZLER, TOBIAS: *Seam Carving*. Seminar Ringvorlesung, HAW Hamburg, 2009. [www.informatik.haw-hamburg.de](http://www.informatik.haw-hamburg.de).
- [LSTS04] LI, YIN, JIAN SUN, CHI-KEUNG TANG und HEUNG-YEUNG SHUM: *Lazy snapping*. ACM Trans. Graph., 23(3):303–308, 2004.
- [SAD<sup>+</sup>06] SANTELLA, ANTHONY, MANEESH AGRAWALA, DOUG DECARLO, DAVID SALESIN und MICHAEL COHEN: *Gaze-based interaction for semi-automatic photo cropping*. In: *CHI '06: Proceedings of the SIGCHI conference on Human Factors in computing systems*, Seiten 771–780, New York, NY, USA, 2006. ACM.
- [Sch] SCHAPIRE, ROBERT E.: *MSRI Workshop on Nonlinear Estimation and Classification, 2002. The Boosting Approach to Machine Learning An Overview*.
- [SLBJ03] SUH, BONGWON, HAIBIN LING, BENJAMIN B. BEDERSON und DAVID W. JACOBS: *Automatic Thumbnail Cropping and Its Effectiveness*. Seiten 95–104. ACM Press, 2003.

- 
- [SWS<sup>+</sup>00] SMEULDERS, ARNOLD W. M., MARCEL WORRING, SIMONE SANTINI, AMARNATH GUPTA und RAMESH JAIN: *Content-Based Image Retrieval at the End of the Early Years*. IEEE Trans. Pattern Anal. Mach. Intell., 22(12):1349–1380, December 2000.