

Effiziente Fußgängererkennung in Städtischen Verkehrsszenen¹

Shanshan Zhang²

Abstract: Fußgänger sind wichtige Teilnehmer im Stadtverkehr-Umgebungen, und damit eine interessante Kategorie von Objekten für autonome Fahrzeuge zu handeln. Automatische Personenerkennung ist allgemein eine wesentliche Aufgabe zum Schutz von Fußgängern aus Kollision. In dieser Arbeit werden wir neue Ansätze untersuchen und entwickeln, für die Interpretation räumlicher und zeitlicher Eigenschaften von Fußgängern, unter drei verschiedenen Aspekten: Form, Wahrnehmung und Bewegung. Wir testen unsere drei Ansätze auf Bild- und Videodaten von Verkehrsszenen im Freien, die wegen der dynamischen und komplexen Hintergründe herausfordernd sind. Die erzielten Ergebnisse zeigen, dass unsere Ansätze state-of-the-Art-Leistung erreichen und übertreffen sowie auch für andere Anwendungen, wie z.B. Indoor-Robotik oder öffentlicher Überwachung eingesetzt werden können.

1 Einführung

Dank der Entwicklung der Informationstechnologie spielen Computer eine immer wichtigere Rolle in unserem alltäglichen Leben. Heutzutage werden Computer erfolgreich und allgemein für die Datenverarbeitung und Kommunikation eingesetzt, um die Handarbeit zu reduzieren und Bequemlichkeit weitgehend zu erweitern. Darüber hinaus erwarten wir Menschen immer intelligenter Computer, die unser Leben in einer aktiven Art und Weise unterstützen könnten. Man kann sich leicht viele Anwendungen von Computern als Assistenten denken: Sie beobachten Kinder zuhause; sie prüfen abnormale Ereignisse an öffentlichen Orten wie Flughäfen; sie warnen Fahrer vor Kollisionen, etc. Um solche Anwendungen zu realisieren, müssen wir Computer sehen und unsere reale Welt interpretieren lassen. Sehen bedeutet, dass sie die visuellen Daten unserer realen Welt interpretieren können. Zu diesem Zweck müssen wir nur die erfassten Bilddaten an die CPUs übertragen, was leicht durch aktuelle Hardware getan werden kann. Interpretieren ist jedoch viel komplizierter. Man kann argumentieren, dass selbst ein Kind leicht die Umgebung erkennen kann. So sollte es noch leichter für einen modernen Computer sein, der eine relativ hohe Rechengeschwindigkeit besitzt. Leider ist dies nicht der Fall. Erstens, unser menschliches visuelles System erhält analog Signale, während die Computer digitale Signale in Form von binären Bits erhalten. Zweitens, jeder Mensch ist einer großen Menge von visuellen Daten ausgesetzt nachdem er oder sie geboren ist, aber solche umfangreiche visuelle Ausbildung ist kaum für einen Rechner zu erreichen. So ist es immer noch ein Rätsel für Informatiker, wie das menschliche visuelle System so effizient die verschiedenen Gegenstände von solcher großen Menge von Daten erkennen kann.

¹ Englischer Titel der Dissertation: "Efficient Pedestrian Detection in Urban Traffic Scenes"

² Max Planck Institut für Informatik, shanshan@mpi-inf.mpg.de

Was wir versuchen zu tun ist, Computern beizubringen, verschiedene Objekte in einer menschenähnlichen Art und Weise zu erkennen. In den letzten Jahrzehnten haben Forscher in den Computer-Vision- und Mustererkennungs-Gemeinden große Anstrengungen gemacht, um zu ermöglichen, Computer die Umgebung durch die Analyse und Interpretation von Bild- bzw. Videodaten verstehen zu lassen. Leider sind die Computer immer noch weit hinter der Leistungsfähigkeit des menschlichen Gehirns bei der Wahrnehmung dieser Analyse und Folgerung, in Bezug auf sowohl Genauigkeit wie auch Effizienz. Es ist noch ein langer Weg für Computer Vision Wissenschaftler zu gehen. Allerdings ist diese Forschung sinnvoll und spannend, weil sie weitgehend die Fähigkeiten von Computern erweitert, so dass sie den Menschen in einer besseren Weise dienen könnten.

Um das letzte Ziel zu erreichen ist die Erkennung von Objekten eine primäre und wichtige Aufgabe, die wesentlichen Elemente für jedes Szenario sind. Die Fähigkeit zur Erkennung von Objekten würde als Vorverarbeitung auf intelligente Vision-System dienen, von dem es erwartet wird, einige bestimmten Personen zu identifizieren, die menschlichen Verhaltensweisen zu erkennen oder auch komplexere gesellschaftliche Ereignisse zu verstehen.

Man kann sich fragen, wie viele Objektkategorien wir in unserer realen Welt haben? Inmitten eines öffentlichen Bilddatenbank, nämlich ImageNet³, werden mehr als 80,000 Objektkategorien nach dem WordNet⁴ Hierarchie definiert. Aber für viele Anwendungen, ist Person der interessanteste Kategorie, weil die Menschen aktiv an gesammelten visuellen Daten beteiligen, einschließlich der privaten Alben und Video-Überwachung. Insbesondere dient das Szenario von Lokalen als eher wichtiger, weil sie in der Regel mit Sicherheit betrifft. Deshalb konzentriert sich unsere Studie auf Erkennung von Menschen an öffentlichen Orten.

Man kann argumentieren, dass andere Sensoren auch die Aufgabe der Erkennung von Menschen schaffen. So haben beispielsweise Lasersensoren gute Leistung für Objektklassifizierung in Outdoor-Umgebungen durch die Analyse von 3D-Punktwolken-Daten erreicht. Ein offensichtlicher Nachteil der Laser-Sensoren ist jedoch der hohe Preis, vor allem, wenn genaue und dichte Punktwolke für reale Anwendungen erforderlich sind. Eine weitere Option können Ultraschall-Sensoren sein, die billig und bequem verwendet werden. Leider sind sie zu empfindlich gegenüber Lärm und können kaum in der Außenanlage Verkehrs Umgebungen eingesetzt werden. Daher betrachten wir Kameras als eine bessere Wahl, weil Kameras von Low-Cost und in der Lage sind, reichhaltige Informationen über die umliegende Umgebung zu erwerben.

In dieser Arbeit beschäftigen wir uns mit dem Problem der Lokalisierung von Fußgängern von Bild- oder Videodaten, die vor allem von einer auf einem fahrenden Fahrzeug durch regelmäßige Verkehrs in einem städtischen Umfeld montierten Kamera aufgenommen wurden. Diese Aufgabe ist aufgrund einer Reihe von Gründen besonders schwierig, die später diskutiert werden.

³ <http://www.image-net.org/>

⁴ <http://wordnet.princeton.edu>

2 Anwendungen

Fußgängererkennung hat in den letzten Jahrzehnten große Aufmerksamkeit in akademischen Gemeinschaften angezogen. Dies weist vor allem auf die verschiedenen Anwendungen in verschiedenen Bereichen auf. Hier listen wir drei gängige Anwendungen:

2.1 Fahrerassistenzsysteme (ADAS)

Dies ist wahrscheinlich eine der wichtigsten und schwierigsten Anwendungen. Fußgänger sind sehr anfällig Teilnehmer im Stadtverkehr. Verkehrsunfällen wurden als eine der wichtigsten Ursachen für Tod und Verletzungen auf der ganzen Welt in einem Bericht der Weltgesundheitsorganisation aus dem Jahre 2004 beschrieben, die schätzungsweise 1,2 Millionen Todesfälle und 50 Millionen Verletzungen pro Jahr aufweisen. Nach den Angaben der Weltbank-Website⁵ machen Fußgänger 65% der Todesopfer aus den 1,170,000 verkehrsbedingten Todesfälle auf der Welt, wovon 35% Kinder sind. In China, Fußgänger und Radfahrer entfielen im Jahr 1994 27% und 23% der Todesopfer, im Vergleich zu jeweils 13% und 2% in den Vereinigten Staaten. Daher ist der Schutz der Fußgänger von großer Bedeutung. Das erste und wichtigste Ziel ist es, dass Fußgänger bereits die Kollision erkennen könnten, wenn sie sich noch in einer sicheren Entfernung sind. Der Zweck der ADAS ist, möglichst frühzeitig die Fahrer vor möglichen gefährlichen Situation zu warnen, so dass sie genügend Zeit zum Bremsen haben.

2.2 Visuelle Überwachung

Closed-Circuit Television (CCTV) Kameras sind häufig zum Zweck der Überwachung an wichtigen öffentlichen Orten wie Flughäfen, Einkaufszentren und Verkehrsknotenpunkten installiert. Heutzutage werden die Kameras verwendet, um Videodaten aufzuzeichnen, die in Festplatten für den späteren Gebrauch aufbewahrt oder durch Sicherheitspersonal überwacht werden. In der Tat ist es unwahrscheinlich, alle Kameras in einer nicht-automatischen Weise zu überwachen. Um die menschliche Arbeit zu verringern haben die Menschen das Ziel, intelligente Überwachungssysteme zu entwerfen, die in der Lage sind, verschiedene Erkennung und Erkennungsaufgaben in einer automatischen Art und Weise zu erreichen. Fußgängererkennung spielt eine wichtige Rolle für solche Systeme und die Positionsinformation von Personen kann als grundlegende Hinweise für die weitere Analyse von abnormalen Ereignissen wie Raub, Diebstahl oder Laufen rote Lichter verwendet werden.

2.3 Mensch-Roboter-Interaktion

In den frühen Tagen der künstlichen Intelligenz werden Roboter in bestimmten Umgebungen entwickelt, in denen die Menschen nur selten angezeigt sind, um schwere, langwierige

⁵ <http://www.worldbank.org/html/fpd/transport/Straßen/safety.htm>

oder gefährliche Aufgaben auszuführen. Aber heute erwarten die Menschen intelligentere Roboter, die Menschen in einem direkteren Weg dienen kann. Zu diesem Zweck ist die Mensch-Roboter-Interaktion eine wichtige Modul zu entwickeln. Der Grundgedanke der Mensch-Roboter-Interaktion ist, Roboter genau die laufenden Menschen rund um sie lokalisieren zu lassen. Eine wesentliche Anwendung kann Roboternavigation, wobei die Ortsinformationen von Menschen verwendet werden können, um Kollision zu vermeiden. Darüber hinaus erwarten die Leute, dass die Roboter einige weiterführende Aufgaben durchführen könnten, z.B. Menschen bei der Bewegung bzw. bei der Lieferung einiger Objekte zu unterstützen, oder auch die Menschen beim Abendessen zu bedienen wie die Kellner in Restaurants. Um diese komplexen Aufgaben zu erfüllen sollten die Roboter in der Lage sein, detaillierte Informationen über umgebende Menschen zu erwerben, wie zum Beispiel Positionen, Größenhaltungen usw.

Unter den obigen drei Anwendungen gilt eine der Fahrerassistenzsysteme (ADAS) als größte Herausforderung sein. Der Grund liegt darin, dass das Außen Szenario gewöhnlich aus mehr komplexen Hintergrund besteht, und die sich bewegende Kamera verursacht signifikante Änderungen zwischen zwei aufeinanderfolgenden Bildrahmen selbst in einem kurzen Zeitfenster. Beide Gründe führen zu bemerkenswerten Störungen für Menschen-Erkennung. Im Gegensatz dazu wird das Szenario die visuellen Überwachung, die Kamera ist statisch und der Hintergrund ändert sich geringfügig über die Zeit; für die Anwendung der Mensch-Roboter-Interaktion ist die Szenario in der Regel weniger dynamisch, weil sich die Roboter in der Regel langsamer als Fahrzeuge bewegen.

Deshalb konzentrieren wir uns auf die Anwendung der Fahrerassistenzsysteme (ADAS). Dennoch sind die in dieser Arbeit entwickelten Technologien nicht auf Verkehrsszenen beschränkt, sondern sie können in anderen Anwendungen direkt oder nach geringen Modifikationen verwendet werden.

3 Herausforderungen

Fußgängererkennung in Verkehrsszenen ist eine anspruchsvolle Aufgabe, nicht nur weil sich die menschliche physische Erscheinungsbild, Kleidung und Körperhaltungen erheblich variieren kann; sondern auch aufgrund der Abweichungen von den realen Umgebungen. Wir stellen kurz die großen Schwankungen in beiden Aspekten vor.

Intra-Class-Variationen: Fußgänger sind hohe Intra-Klasse-Variante, die deutlich signifikanter als einige andere Objektkategorien wie Autos oder Gebäuden sind. Diese Intra-Klasse-Schwankungen machen es unmöglich, die Fußgänger von anderen Objekten von einem einzigen Cue, wie Farbe oder Form zu unterscheiden. Im Folgenden beschreiben wir Intra-Klasse-Varianten in Bezug auf Aussehen, Kleidung Entscheidungen undhaltungen.

- Aussehen: Menschen können aufgrund genetischer Eigenschaften ganz unterschiedlich voneinander sein, die verschiedenen Hautfarben, Haarfarben, Augenfarben, Figuren haben. Man kann auch sein ursprüngliches Aussehen mit Make-up-Fähigkei-

ten, einschließlich das Ändern von Frisuren ändern. Darüber hinaus zeigt man in der Regel verschiedene Auftritte in verschiedenen Altersstufen an.

- **Einsätze von Kleidung:** Man kann Kleidung mit einer breiten Vielzahl von Farben, Texturen und Designs wählen. Manchmal tragen die Menschen auch verschiedene Accessoires wie Schmuck oder Handtaschen. Alle oben genannten Faktoren führen zu wenigen einheitlichen Farben oder Texturen der inneren Bereiche des menschlichen Körpers. Dementsprechend wird es darauf hingewiesen, dass innere Bereiche unzuverlässige Einsätze sind.
- **Haltungen:** Eine weitere Quelle von Variationen stammt aus Haltungen der menschlichen Körper Exponate. Fußgänger können zu Fuß über die Straße laufen, an der Ampel stehen oder an eine Wand lehnen. Da die menschlichen Körper nicht starr ist, dass eine Vielzahl von Stellungen wahrscheinlich erscheinen. Jede Körperhaltung zeigt eine identische Form, so dass die Leute mit unterschiedlichen Haltungen ganz verschieden voneinander aussehen. Das heißt, wir müssen vorsichtig bei Formdeskriptoren für Menschen-Erkennung sein.

Umwelt Variationen: Städtische Verkehrsumgebung ist sehr komplex und wenig eingeschränkt, die in der Regel aus verschiedenen Objektkategorien besteht, z.B. Fahrzeugen, Gebäuden, Pflanzen, Tieren und Menschen. Darüber hinaus sind die Standorte der Fußgänger in Bezug auf der Kamera unterschiedliche Standpunkten, Bedingungen der Okklusion oder Schuppen, und deswegen sind sie auch ein wichtiger Faktor, der bestimmt, wie Fußgänger in den Bild- oder Videodaten aussehen.

- **Dynamische Hintergrund:** Wenn die Kamera sich bewegt mit dem fahrenden Fahrzeug, wird der Hintergrund dynamisch. Im Vergleich zu einem statischen Szenario, in dem der Hintergrund modelliert werden kann und sich bewegende Objekte durch Hintergrundabzug gefunden werden können, ist der dynamische Hintergrund anspruchsvoller, weil jedes Objekt seine Lage und das Aussehen zwischen zwei aufeinanderfolgenden Frames auch in einem eher kleinen Zeitfenster ändern kann. Es wird vorgeschlagen, die Auswirkungen der Kamerabewegung durch die Verwendung affine Theorien zu beseitigen, dennoch ist es eine sehr schwierige Aufgabe, vor allem wenn das Fahrzeug mit hoher Geschwindigkeit fährt. Zu diesem Zweck ist es ziemlich schwierig, Verwendung von Bewegungsinformationen zur Fußgängererkennung zu machen.
- **Viewpoints:** Grundsätzlich werden Standpunkte durch die relative Lage zwischen Kamera und beobachteten Menschen bestimmt. In den meisten Fällen weisen Objekte unterschiedliche Erscheinungsbilder auf aus verschiedenen Blickwinkeln. Zum Beispiel sehen die Autos viel breiter aus von der linken oder rechten Seite im Vergleich zu den von vorne oder von hinten aussehen. Glücklicherweise ist das Seitenverhältnis der Fußgänger weniger Gesichtspunkt relevant als Autos. Jedoch führen unterschiedliche Standpunkte immer zu verschiedenen Auftritten. Beispielsweise sieht man größere Winkel zwischen zwei Schenkeln des laufenden Fußgänger von der Seitenansicht als die von der Vorder- oder Rückansicht. Außerdem sehen die Schulter größer aus von der Vorder- oder Rückansicht als die in der Seitenansicht.

- **Okklusion:** In den Stadtverkehr-Szenen ist Okklusion sehr häufig. Fußgänger können durch andere Objekte oder sogar benachbarten Fußgänger in einer überfüllten Szene verdeckt werden. In der ersten Situation, die fehlenden Körperteile in der Regel ändern die Gesamterscheinung. Für diejenigen Klassifikatoren können die unverschlossenen Fußgänger als eine nicht-Fußgänger Objekt erkannt werden, was zu einer hohen Miss-Raten dient. Im zweiten Fall ist es schwierig, zu schließen, welcher Körperteil zu welcher Person gehört, wenn mehrere Fußgänger hochklappen. Manchmal wird eine Gruppe von Fußgängern als einziger erfasst, während manchmal mehrere Hypothesen vorgenommen werden.
- **Skalen:** Skalierung des Objekts hängt hauptsächlich von der Entfernung zwischen dem betrachteten Objekt und der Kamera ab, unter der Annahme, dass wir die individuelle Differenz in der Höhe ignorieren, die nicht signifikant für Fußgänger ist. Auf diese Weise zeigen die Fußgänger, die weit von der Kamera entfernt sind, wenige Pixel als die näher von der Kamera entfernte in Bilddaten. Die Herausforderungen sind zweifach. Erstens können einige wichtige Hinweise, wie Verbindungen zwischen Körperteilen vage werden, wenn die Anzahl der Pixel unter einer bestimmten Schwelle geht. Manchmal können wir nur eine Silhouette sehen, die noch schwieriger für Menschen zu erkennen und leicht als ein Baum an einigen Fällen anerkannt ist. Zweitens müssen wir Skalen für Schiebefenster Ansätze durchsuchen, damit die Fußgänger von unterschiedlichen Skalen gefunden werden können. Diese sich nahe an der Kamera befindenen Fußgänger sind fast zehnmal größer als die sich entfernten stehenden Fußgänger.

4 Beiträge des Thesis

Diese Arbeit untersucht das herausfordernde Problem der Fußgängererkennung im Stadtverkehr Szenen und schlägt vor, neue Verfahren zur Interpretation der Merkmale der Fußgänger in drei verschiedenen Aspekten: Form, Wahrnehmung und Bewegung, um eine effiziente Fußgängererkennung zu ermöglichen.

Form spielt eine wichtige Rolle für die Objekterkennung. Auf der einen Seite zeigen Objekten aus verschiedenen Kategorien verschiedene Formen auf den Bildern; auf der anderen Seite teilen verschiedene Objektinstanzen in der gleichen Kategorie in der Regel einige gemeinsamen Formen.

Durch die Beobachtung einer Vielzahl von Fußgänger Bilder haben wir gefunden, dass für Fußgänger die gemeinsame Form des aufrechten menschlichen Körper ein deutliches Unterscheidungsmerkmal ist, die nicht nur das Problem der Fußgängererkennung viel einfacher als Erkennung von allgemeinen Personenerkennung gemacht hat, sondern auch unterscheiden sich die Fußgänger von anderen Objekten im Stadtverkehr Szenen. Daher wollen wir herausfinden, wie diese Eigenschaft für eine effektivere Fußgängererkennung nutzen kann.

Zu erreichen oder sogar für menschenähnlichen Anerkennung ist das endgültige Ziel für die intelligenten Vision-Systeme. Mit dem Ziel der Gestaltung einer erfolgreichen Fuß-

gänger Detektor, der in der Lage ist, Fußgänger so genau und schnell zu finden wie den Menschen, lohnt es sich zu untersuchen, wie menschliche visuelle Systeme visuelle Daten verarbeiten und Gegenstände lokalisieren, für die sie interessiert sind. Auf diese Weise versuchen wir, ein Design- Kognition System zur Fußgängererkennung zu entwerfen.

Durch die Untersuchung der Mechanismen der Kognition Systemen haben wir die Mitte-Surround-Mechanismus für die hervorstechende Objekterkennung als sehr attraktiv gefunden. Die hervorstechenden Aufgaben werden entsprechend als Fußgänger für eine Fußgängererkennungssystem definiert, in dem unter Verwendung des Mitte-Surround Mechanismus für die Feature-Design ein direkter Weg ist, um eine top-down-saliency System zu emulieren.

Bewegung ist auch ein wichtiger Cue, jedoch ist sie noch nicht in großem Umfang für Fußgängererkennung. Im Gegensatz zu der Form Cue, die räumliche Informationen darstellt beschreibt Bewegung die zeitliche Information in aufeinanderfolgenden Einzelbildern. Wir heben drei Fragen: (1) Zu welchem Zweck können wir Bewegungsinformation benutzen? (2) Wie kann man zeitliche Informationen in einer angemessenen Art und Weise interpretieren? (3) Erzeugt Bewegungsinformation Verbesserungen während sie integriert mit räumlichen Informationen? Um diese Fragen zu beantworten, fanden wir heraus, verschiedene inter-Klasse und intra-Klassenmerkmale für Fußgänger durch die Beobachtung einer großen Anzahl von Bewegungskarten, optischen Strömungsvektoren darzustellen. Diese Erkenntnisse inspirieren uns, ein spezifisches Problem von sich bewegender Fußgängererkennung zu lösen.

Im Folgenden werden wir kurz die wichtigsten Beiträge dieser Arbeit in einer technischen Weise vorstellen.

- In Kapitel 3 [ZBC14, ZBC15] schlagen wir ein statistisches Formmodell für den aufrechten menschlichen Körper vor. Dann entwerfen wir abgestimmte Haar-ähnliche Funktionen auf diese Formmodell, so dass die besondere Form-Struktur in Bezug auf der lokalen Differenz repräsentieren. Um robuster Deskriptoren für die lokale Unterschiede zu erzeugen, betrachten wir mehrere Bild-Kanälen, wobei die Farben, Gradientengröße und Histogramme von orientierten Gradienten miteinbezogen sind. Wir haben auch eine ternäre Modalität als Zusatz eingeführt zur traditionellen binären Modalität, um komplizierter geometrischer Konfigurationen darzustellen. Diese informierte Funktionen vermeiden erschöpfende Suche über alle möglichen Konfigurationen und weder stützen sich auf eine Zufallsstichprobe von einem rechteckigen Merkmalsraum und markieren damit einen Mittelweg. Die vorgestellten experimentellen Ergebnisse zeigen, dass unsere Eigenschaften erreichen sowie übertreffen state-of-the-art Performance und robust für Verstopfungen sind. Darüber hinaus erfordern unsere Funktionen wenige Speicher und Rechenzeit für die Ausbildung als kürzlich vorgeschlagenen Wettbewerbsdetectoren und es wird erwartet, sie Echtzeit-Performance mit GPU-Berechnung erreichen können. Das Flussdiagramm wird gezeigt in Abb. 1.
- In Kapitel 4 [Zh14, Zh15a] schlagen wir vor, lokale Kontrastmerkmale durch die Mitte-Surround-Mechanismus in menschlichen visuellen Systemen zu motivieren

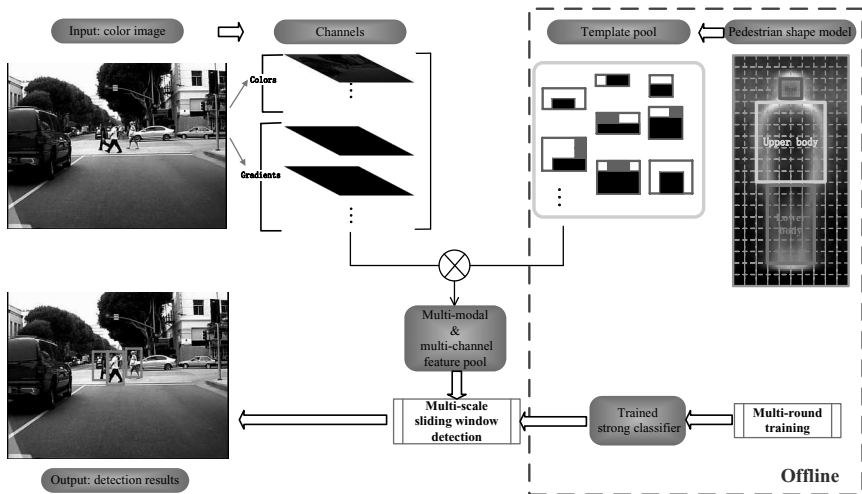


Abb. 1: Übersicht über die Fußgängererkennung basierend auf informierte Haar-ähnliche Funktionen.

und sie auf die Erscheinung von Fußgängern reagieren zu lassen. Im Gegensatz zu früheren Verfahren, bei denen Kontrastpixelweise berechnet werden, betrachten wir Kontrastwerte für jeden Bereich, die durch eine statistische Deskriptor statt durch Bildkanal Werte direkt dargestellt wird. Um reichere Informationen zu den lokalen Unterschieden zu integrieren, führen wir multi-direktionale Gegensatz Vektoren ein, die umgebenden Zellen einzeln und nicht als ein Ganzes zu behandeln. Eine weitere wichtige Aufgabe ist es, eine angemessene Abstandsmessung für die Region Deskriptoren zu wählen. Wir machen einen umfassenden Vergleich der Descriptor Streckenkombinationen und finden Sie heraus die optimale von allen Experimenten. Ähnlich wie bei den meisten saliency Detektionssystemen bauen wir eine Kontrastpyramide, was grob-zu-fein lokalen Unterschied durch Variation der Zellgröße darstellt. Das Flussdiagramm wird gezeigt in Abb. 2.

- In Kapitel 5 [Zh13, Zh15b] konzentrieren wir uns auf eine weitere spezifische Aufgabe der Erfassung von bewegenden Fußgänger, die eine interessante Untergruppe für die Anwendung der Adass ist. Bewegungsinformation wird auf zwei verschiedene Arten verwendet. Einerseits führen wir Graphen basierte Segmentierung optischen Fluss Karten zweidimensionalen durch, um Bereiche von Interesse (ROIs) und sich bewegende Objekten durch Blob-Analyse auszuwählen. Auf der anderen Seite entwerfen wir Bewegung Selbst Unterschied kennzeichnet, da wir deutliche Größenordnung Karten bewegter Fußgänger von anderen bewegten Objekten beobachten. Schließlich, um verschiedene Kategorien von Funktionen in einem Lern Rahmen zu integrieren, führen wir ein Zwei-Schicht-System für zuverlässigere Klassifizierung. Das Flussdiagramm wird gezeigt in Abb. 3.

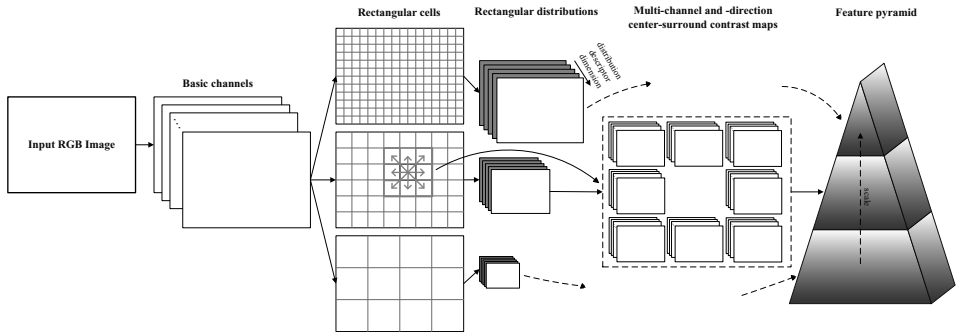


Abb. 2: Flussdiagramm der Mitte-Surround-Merkmalsextraktion.

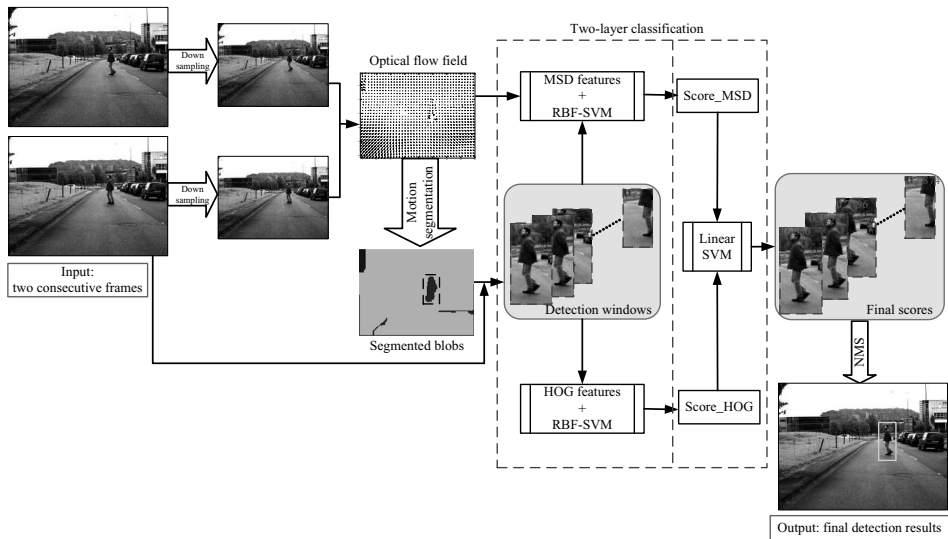


Abb. 3: Flussdiagramm der Bewegung basierte Fußgängererkennung bewegen.

Unsere Ansätze erreichen und übertreffen state-of-the-art Ergebnisse aus Experimenten auf verschiedenen Datensätzen. Im Folgenden fassen wir einige Lehren, die wir aus unserer Untersuchung gelernt haben. Diese Notizen können lehrreich für die zukünftige Forschung sein.

- Vorkenntnisse ist mächtiger als wir dachten. In Kapitel 3 erhalten wir überraschend bessere Ergebnisse als die Baseline-Detektor unter Ausnutzung der Vorkenntnisse für Spieldesign. In der Tat können wir Vorkenntnisse, um die Daten, die wir brauchen, handhaben und damit Verbesserung der Fähigkeit zur Erkennung besser verstehen. Daher ist es sinnvoll zu untersuchen, wie Vorwissen umfassender in der Zukunft zu verwenden.
- Mechanismen der menschlichen visuellen Systeme sind hilfreich für die Gestaltung effektiver intelligenter Vision-Systeme. Das haben Forscher schon vor vielen Jah-

ren erkannt, aber es ist immer noch eine offene Frage, wie diese Mechanismen mit den bestehenden Computer-Vision und Mustererkennungstechniken in angemessener Weise zu integrieren.

- Eigenschaften gegenüber Lernmethoden. Es gibt eine Debatte darüber, welche für eine erfolgreiche Fußgängerdetektor wichtiger ist: Merkmale oder Lernmethoden. In dieser Arbeit arbeiten wir vor allem auf Feature-Design, und hat gefunden, dass aufgrund unserer Verfahren sorgfältig entworfene Funktionen solche recht komplexe Lerntechniken übertreffen. Dieser Erfolg zeigt, dass die Gestaltung der Merkmale die Statistik der angegebenen Daten einzuhalten eine vielversprechende Richtung ist und es voraussichtlich noch bessere Leistung erzielen wird.

Literaturverzeichnis

- [ZBC14] Zhang, S.; Bauckhage, C.; Cremers, A. B.: Informed Haar-like Features Improve Pedestrian Detection. In: Computer Vision and Pattern Recognition (CVPR), 2014 IEEE Conference on. 2014.
- [ZBC15] Zhang, S.; Bauckhage, C.; Cremers, A. B.: Efficient Pedestrian Detection via Rectangular Features Based on a Statistical Shape Model. *IEEE Transactions on Intelligent Transportation Systems*, 16(2):763–775, 2015.
- [Zh13] Zhang, S.; Bauckhage, C.; Klein, D. A.; Cremers, A. B.: Moving pedestrian detection based on motion segmentation. In: Robot Vision (WORV), 2013 IEEE Workshop on. 2013.
- [Zh14] Zhang, S.; Klein, D. A.; Bauckhage, C.; Cremers, A. B.: Center-Surround Contrast Features for Pedestrian Detection. In: Pattern Recognition (ICPR), 2014 22nd International Conference on. 2014.
- [Zh15a] Zhang, S.; Bauckhage, C.; Klein, D. A.; Cremers, A. B.: Exploring Human Vision Driven Features for Pedestrian Detection. *IEEE Transactions on Circuits and Systems for Video Technology*, 25(10):1709–1720, 2015.
- [Zh15b] Zhang, S.; Klein, D. A.; Bauckhage, C.; Cremers, A. B.: Fast Moving Pedestrian Detection Based on Motion Segmentation and New Motion Features. *Multimedia Tools and Applications*, 2015.



Shanshan Zhang erhielt den Master in Signal- und Informationsverarbeitung von der Tongji-Universität im Jahr 2011, und der Promotion in Informatik von der Universität Bonn im Jahr 2015. Derzeit arbeitet sie als Postdoc -Forscher an der Abteilung für Computer Vision und Multimodal Computing, Max Planck- Institut für Informatik, Saarbrücken. Sie war Gastforscher in National Institute of Informatics, Japan, im Jahr 2010. Ihre Forschungsschwerpunkte sind Computer Vision und Mustererkennung und ihre Anwendungen für fahrerlose Fahrzeuge.