

Counting People from Multiple Cameras

Vera Kettner, Ramin Zabih

The IEEE International Conference on Multimedia Computing and Systems, 1999

Johann Engelbrecht

Universität Bielefeld
Seminar Visuelle Überwachung SS 2006

Review

Einleitung

In dem Paper stellen die Autoren ein System vor, mit dem man voneinander verschiedene bewegliche Objekte in einer geschlossenen Umgebung anhand von untereinander interagierenden Kameras erfassen kann. Die Neuheit liegt in der Verwendung von nicht überschneidenden Kameras.

Der Ansatz zur Lösung dieser Aufgabe beruht auf der Bewegungsverfolgung von agilen Objekten. Sie ist ein Hauptaspekt der visuellen Überwachung und oft die Basis für viele Überwachungssysteme.

Relevanz zum Seminar und praktischer Einsatz

In dem Paper wird beschrieben, dass das vorgestellte System flexibel ist und nicht nur auf das Zählen von Personen beschränkt werden muss. Auch andere Objekte können wahrgenommen und die Richtung des überwachten Objekts ermittelt werden.

Durch die Erfassung der Menge an Fahrzeugen und der Aufzeichnung der jeweiligen Fahrtrichtung in Gebieten mit ausgelastetem Straßenverkehr ist zum Beispiel eine bessere Planung von Ampelphasen im Straßenverkehr möglich. Mit Hilfe einer Auswertung der erhaltenen Daten kann der Verkehrsfluss durch eine neue Ampelkonfiguration verbessert werden. Abschließend kann eventuell durch Einsatz des Systems die Bildung von Staus gemindert werden.

Zusammenfassung

In den letzten Jahren ist die Popularität von digitalem Equipment stark gestiegen. Es werden viele neue Methoden benötigt um digitale Daten verarbeiten und nutzen zu können. In diesem Fall wurde ein System entwickelt mit dem man unter anderem verschiedene bewegliche Objekte erkennen und zählen kann. Die Überwachungsmethode muss in der Lage sein einzelne Personen von einander unterscheiden zu können. Aufnahmen von den gleichen Personen in verschiedenen Positionen müssen wahrgenommen und nur einer Person zugeordnet werden. Es ist falsch jede Person in jeder Videosequenz zu zählen, weil manche Personen mehrmals gezählt werden würden wenn sie öfters in Bereiche treten.

Es wurden vier Kameras mit verschiedenen Blickwinkeln aufgestellt. Im Gegensatz zu anderen ähnlichen Projekten wurde bei der Einrichtung beachtet, dass die Winkel sich nicht überschneiden. Bei Installationen mit überschneidenden Kameraeinstellungen kann das Gesamtbild als eine einzige Aufnahme einer Kamera gesehen werden. Aus diesem Grund ist es für ein Programm schwer die Unterscheidung zu treffen wieviele Personen in den Aufnahmen zu sehen sind und welchen Weg sie wählen. Multiple Kamerasysteme decken mehrere räumliche Bereiche ab und sind daher im Vorteil gegenüber einfachen Systemen die nur eine bestimmte Fläche überwachen können. Aufgrund von begrenztem Speichervolumen kann diese Aufgabe einen Menschen leicht überfordern und deswegen wird ein automatisches Erkennungsverfahren benötigt.

Die vorgestellte Methode startet mit dem erstmaligem Auftreten einer bestimmten Person. Das System unterteilt die aufgezeichneten Bilder in Bereiche indem es die Pixel untereinander vergleicht. Bildpunkte die sich ähnlich sehen und sich gleichermaßen verhalten werden in gleiche Regionen segmentiert. Das sogenannte „background subtraction scheme“ hat den Effekt, dass Pixel mit einem leichten Schatten oder minimalen Reflexionen zu den Hintergrundpixeln segmentiert

werden. Die statischen Regionen werden von dem Bild entfernt und man erhält nur die Pixel des beweglichen Objekts. Die Werte der einzelnen Bildpunkte für den Farbton, die Sättigung und die Helligkeit werden gespeichert und mit anderen Werten verglichen. Falls eine andere Aufzeichnung mit ähnlichen Werten gefunden wird, gruppiert das System die Werte und zählt nur eine Person. Man erhält aus den Aufzeichnungen ein statistisches Model der Pixelwerte für jede Person mit dem man verschiedene Personen voneinander unterscheiden kann.

Ein Problem bei der Überwachung mit mehreren Kameras entsteht, wenn die installierten Geräte verschiedene Blickwinkel haben, unterschiedliche Lichtverhältnisse herrschen oder Teile der Objekte verdeckt werden. Um jene Schwierigkeiten zu umgehen wurde die Lichtintensität jeder Kamera an eine einzelne angepasst und ein äußeres Erscheinungsbild der Person gewählt, dass bei derartigen Änderungen stabil bleibt.

Bei dem Experiment wurde der Farbraum so ausgesucht, dass feine Farbunterschiede bei Kleidungsstücken in beige oder grau erfasst werden können, weil diese Farben sehr beliebt sind, während bei unbeliebteren Farben wie rot die Messungen ungenauer durchgeführt wurden. So entstand eine effizientere Lösung.

Durch das Ausnutzen von Wissen über die räumliche Struktur und die durchschnittliche Geschwindigkeit des Ganges eines Menschen kann der Computer Schlussfolgerungen ziehen. Er kann die Möglichkeiten einschränken an welchem Ort und um welche Zeit ein Objekt als nächstes gesichtet werden wird und diese Kenntnisse helfen ihm bei der Verfolgung und Unterscheidung zwischen mehreren Personen. Dadurch werden Verwechslungen oder doppelte Zählungen nur auftreten falls das Objekt ähnliche Messdaten produziert hat, in die gleiche Richtung geht und zur gleichen Zeit von einem Kamerawinkel in einen anderen übergeht. Diese drei Aspekte sollen die Fehlerrate reduzieren. Die Überwachungsaufnahmen werden in zeitlicher Reihenfolge gespeichert. Sie erlauben weitere Vergleiche, wodurch die Identifikation von Objekten eine Effizienzsteigerung erfährt.

Um das System auszuwerten wurde ein Experiment durchgeführt. Es hatte die Dauer von 8 Minuten. Die Anlage bestand aus 4 Überwachungskameras und wurde in und um das Labor installiert. Bei dem Experiment gingen zwei Objekte so oft wie es ihnen möglich war durch das aufgebaute System. Sie haben sich mehrmals umgezogen um verschiedene Personen zu simulieren. Es wurden 28 Sammlungen von Bildausschnitten von 14 überwachten Objekten erfasst, wobei zwei Intervalle von Bildausschnitten falsch interpretiert wurden. So entstand das Resultat dass 12 von insgesamt 14 überwachten Objekten erkannt und gezählt wurden.

Diskussion

In dem Paper wird deutlich dass die Autoren sich der Stärken und Schwächen ihres Systems bewusst sind. Sie beschreiben wo ihr System seine Vorteile ausspielen kann und sind im Großen und Ganzen ehrlich in Bezug auf die Fähigkeiten. Ich bewerte diese Aspekte positiv weil es keinem was bringt falsche Tatsachen vorzutäuschen. Wenn ein System irgendwo zum Einsatz kommt wird es sicherlich vorher getestet und spätestens dann werden Schwächen und Stärken sichtbar. So kann der Benutzer selber entscheiden ob er es in seinem Fall einsetzen möchte oder nicht.

Weiterhin hat mir gefallen, dass im Paper alles recht deutlich erklärt und dargestellt wird. Es wird zwar nicht sehr tief in die Materie eingedrungen, doch für einen allgemeinen Überblick reicht es vollkommen aus.

Trotzdem sind mir auch eine Menge negativer Aspekte an dem Paper aufgefallen, die mir nicht gefallen haben. Ich fand die Darstellung des Systems weniger aussagekräftig. Es werden Abbildungen benutzt die nur wenig darstellen. Als Beispiel möchte ich Abbildung 3 erläutern. Auf ihr ist nur zu erkennen wann Personen von welcher Kamera erfasst worden sind. Ich finde sie macht nur wenig Sinn, weil sie im Text auch nur kurz eingebunden wird.

Nach der Bearbeitung der Darstellung des Systems blieben bei mir ein paar Fragen offen. Es wurde darauf hingewiesen das bei der Erkennung von Objekten leichte Schattierungen und geringfügige Reflexionen aus dem Bild gelöscht wurden. Mir stellt sich dabei aber die Frage was mit den Schatten und Reflexionen geschieht, die sehr dominant sind. Werden diese immer mit zu dem Objekt gezählt. In diesem Fall würde die Effizienz eine Verschlechterung erfahren, weil bei jedem Vergleich mehr gerechnet werden müsste. Falls der Fall eintritt dass der Schatten ein

weiteres Objekt für das System darstellt, würde die Fehlerrate in lichtintensiven Räumen steigen. Außerdem wurde nicht deutlich genug erläutert wie das System mit mehreren Objekten zur selben Zeit zurecht kommt. In der Wirklichkeit ist die Wahrscheinlichkeit hoch, dass mehrere Personen gleichzeitig im Bild zu sehen sind. Weiter möchte ich die räumliche Struktur, in der getestet wurde, bemängeln. Sie ist übersichtlich aufgebaut, so hat das beobachtete Objekt wenige Möglichkeiten um unvorhersehbare Aktionen simulieren zu können. Der Farbraum des Systems wurde speziell für Farben wie Grau oder Beige fein justiert, weil es sehr populäre Farben sind. Meiner Meinung nach darf man das nicht verallgemeinern, weil es immer auf den Standort des Systems und die Jahreszeit ankommt welche Farben dort überwiegend getragen werden. Zum Beispiel wird im Sommer öfter sehr bunte Kleidung getragen im Gegensatz zum Winter wo unauffälligere Kleidung den Besitzer schmückt.

Die ausgeführte experimentelle Auswertung zur Ermittlung wie robust das System arbeitet bewerte ich ebenfalls negativ. Das Experiment dauerte nur 8 Minuten und es wurde nur mit zwei Personen durchgeführt. Diese haben sich oft umgezogen und damit verschiedene Menschen simuliert. So konnten trotzdem höchstens zwei Objekte gleichzeitig visuell verfolgt werden was im Alltag eventuell nicht eintreffen wird. Das Ergebnis der Tests lässt ebenfalls zu wünschen übrig. Das System erkannte 12 von 14 verschiedenen Objekten und hat damit eine Trefferquote von ca. 85%. Meiner Meinung nach ist das Ergebnis nicht zufriedenstellend, mit der Hinsicht auf die Dauer des Experiments. Aus einer so kurzen Periode können keine repräsentativen Daten entstehen.

Alternativen

In dem Paper werden Projekte genannt die dem vorgestellten System ähneln. Es wird deutlich gemacht das sie vergleichsweise ein besseres System erstellt haben. Diese Aussage kann ich ohne Wissen über die anderen Systeme nicht beurteilen. Den Vergleich mit dem System von Huang und Russell [6] finde ich aber merkwürdig, da die Aufgabe darin bestand Menschen mit Hilfe von multiplen Kameras zu erfassen. Es wurde in einer geschlossenen Umgebung getestet und dem entsprechend konfiguriert. Ich kann den Vergleich mit einem System, das Autos auf einem Highway zählt, nicht nachvollziehen, weil die Konfiguration des menschenzählenden Systems nicht dafür konzipiert ist. Es wurde ausgesagt, dass diese Überwachung vor allem sehr gut in geschlossenen Räumen angewendet werden kann. Somit stellt diese Aussage einen Widerspruch zum Vergleich mit dem System von Huang und Russell dar, da dieses im Freien getestet wurde.

Persönliches Fazit

Mein Fazit lautet: das vorgestellte System genügt nur dann den Ansprüchen, wenn man dessen Einsatz über einen langen Zeitraum hinweg durchführen will. Außerdem wäre es von Vorteil wenn der Einsatzort und Einsatzzweck mit dem getesteten möglichst übereinstimmt, weil sonst sehr viele Konfigurationen verändert werden müssen, wie zum Beispiel die Intensitätsnormierung der Kameras und die Farbraumanpassung. Die räumliche Struktur muss ebenfalls besonders gewählt werden. Alles muss auf die gegebene Situation eingestellt werden. Das kostet sehr viel Zeit und ist mit hohen Kosten verbunden. In dem Paper wird beispielsweise erläutert wo das System eingesetzt werden kann. Es wird der Aufgabenbereich Autowiedererkennung genannt. Sie wählen diesen aufgrund von vorherrschenden Bedingungen. Autos sind statisch und können nicht so individuell reagieren wie Menschen. Es wird aber auch nicht bedacht was für Konditionen bei der Autowiedererkennung zwingend vorherrschen müssen. Dort gibt es wechselnde Lichtverhältnisse, Reflexionen von Lackierungen, also eine komplett andere Struktur. Das bedeutet dass das System bei einem neuen Einsatz noch einmal neu strukturiert und konfiguriert werden müsste.

Insgesamt ist zu sagen, dass das System seine Vorteile hat, wenn es in einer Umgebung installiert wird, die der getesteten sehr ähnlich ist. Die Anpassung an andere Aufgaben und Standorte wäre sehr aufwendig. Man müsste wahrscheinlich andere Lösungsansätze finden um die Unterschiede zwischen den Aufgabengebieten zu kompensieren. Das System wurde mit der starken Hinsicht auf die Testumgebung und -situation erstellt und kann nur Einsatz in einer geschlossenen Umgebung finden.

Literatur

[6] Tim Huang, Stuart Russell, Object identification: a Bayesian analysis with application to traffic surveillance. *Artificial Intelligence*, 103: 1-17, 1998.