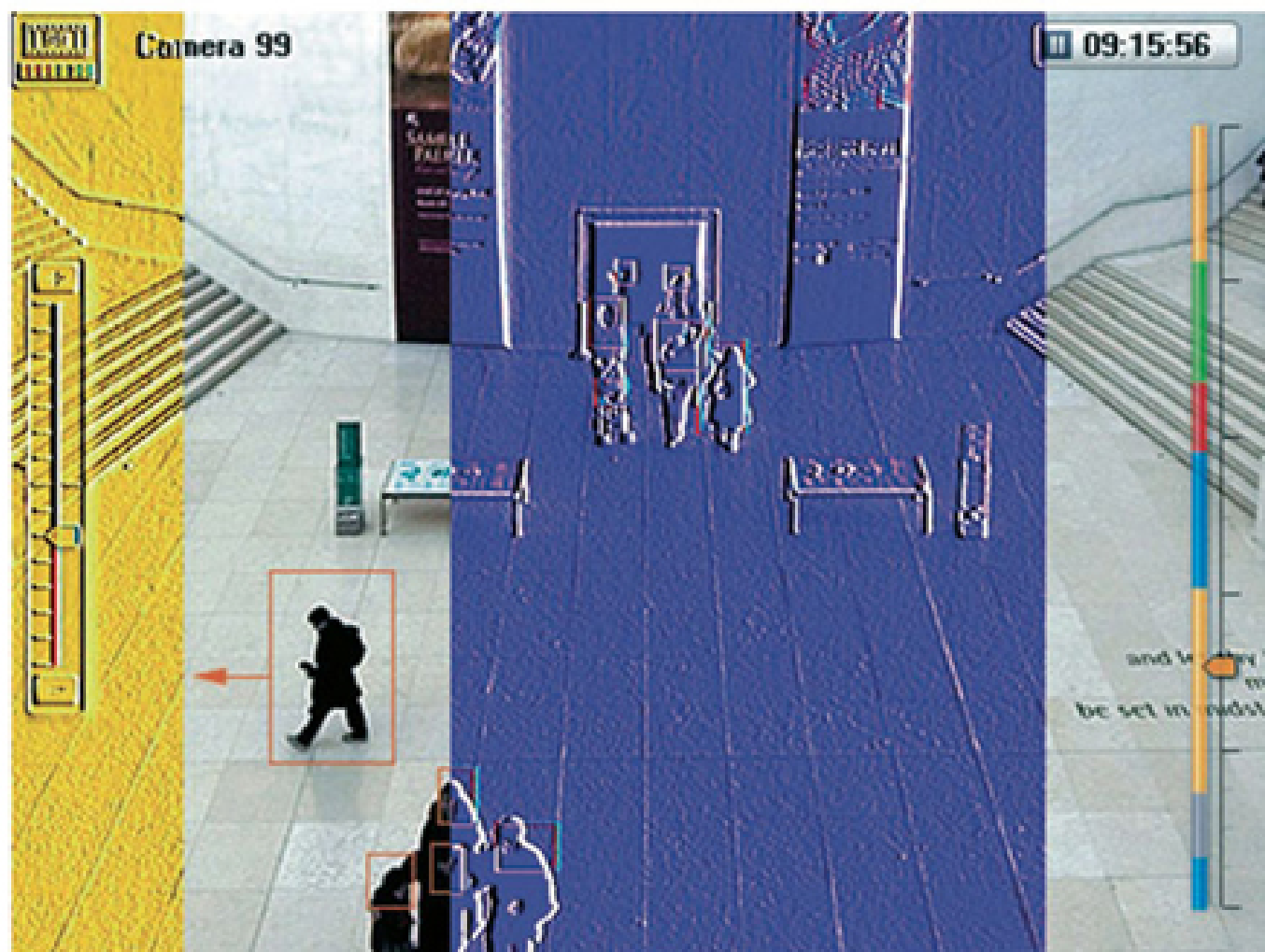


Torsten Anstädt, Ivo Keller
und Harald Lutz

 WILEY-VCH

Intelligente Videoanalyse

Handbuch für die Praxis

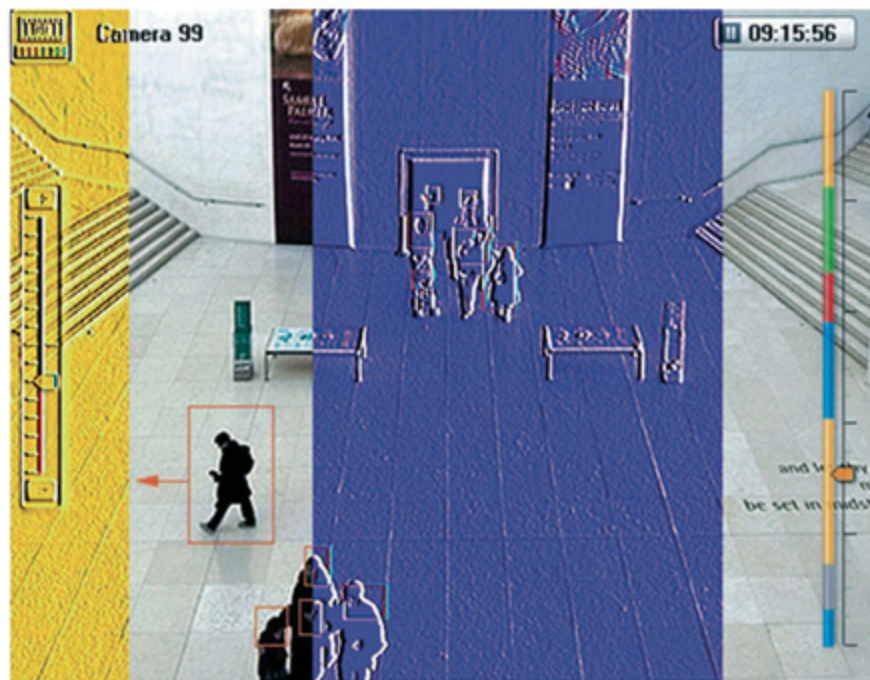


Torsten Anstädt, Ivo Keller
und Harald Lutz

WILEY-VCH

Intelligente Videoanalyse

Handbuch für die Praxis



Torsten Anstädt, Ivo Keller und Harald Lutz

Intelligente Videoanalyse: Handbuch für die Praxis

Torsten Anstädt, Ivo Keller und Harald Lutz

Intelligente Videoanalyse

Handbuch für die Praxis



**WILEY-
VCH**

WILEY-VCH Verlag GmbH & Co. KGaA

Autoren

Torsten Anstädt

Wiesbaden, Deutschland anstaedt@web.de

Dr. Ivo Keller

Potsdam, Deutschland ivo.keller@gmx.net

Dipl.-Ing. Harald Lutz

Worfelden, Deutschland hlutz@worfelden.de

Titelbild

Great Hall in the British Museum Video by AxxonSoft GmbH

1. Auflage 2010

Alle Bücher von Wiley-VCH werden sorgfältig erarbeitet. Dennoch übernehmen Autoren, Herausgeber und Verlag in keinem Fall, einschließlich des vorliegenden Werkes, für die Richtigkeit von Angaben, Hinweisen und Ratschlägen sowie für eventuelle Druckfehler irgendeine Haftung

Bibliografische Information

der Deutschen Nationalbibliothek

Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über <http://dnb.d-nb.de> abrufbar.

© 2010 WILEY-VCH Verlag GmbH & Co. KGaA, Boschstr. 12, 69469 Weinheim

Alle Rechte, insbesondere die der Übersetzung in andere Sprachen, vorbehalten. Kein Teil dieses Buches darf ohne schriftliche Genehmigung des Verlages in irgendeiner Form – durch Photokopie, Mikroverfilmung oder irgendein anderes Verfahren – reproduziert oder in eine von Maschinen, insbesondere von Datenverarbeitungsmaschinen, verwendbare Sprache übertragen oder übersetzt werden.

Die Wiedergabe von Warenbezeichnungen, Handelsnamen oder sonstigen Kennzeichen in diesem Buch berechtigt nicht zu der Annahme, dass diese von jedermann frei benutzt werden dürfen. Vielmehr kann es sich auch dann um eingetragene Warenzeichen oder sonstige gesetzlich geschützte Kennzeichen handeln, wenn sie nicht eigens als solche markiert sind.

ISBN 978-3-527-40976-1

Inhaltsverzeichnis

Cover

Half Title Page

Title Page

Copyright

Vorwort

**Wir danken recht herzlich für die
Unterstützung.**

**1 Historie und wissenschaftliche
Perspektive**

**2 Algorithmen der intelligenten
Videoanalyse**

**2.1 Klassische Bewegungserkennung
Motion Detection**

2.2 Personenmodelle

2.3 Vordergrund-/Hintergrundanalyse

2.4 Maschinelles Lernen

2.5 Mustererkennung

**2.6 Zweidimensionale Abbildung einer
dreidimensionalen Welt**

2.7 Objekttracking

2.8 Szeneninterpretation

3 Einsatzgebiete der Videoanalyse

3.1 Intelligente Kamera versus PC-gestützte Auswertung

3.2 Infrarot-Licht, atmosphärische Fenster, Eigenstrahlung - Sehen in dunklen Welten

3.3 Terahertz-Wellen - Sehen zwischen Licht und Radar

3.4 Motion Tracking

3.5 Klassifikation

3.6 Perimeterschutz

3.7 Gesichtsdetektion - auf die richtige Größe kommt es an

3.8 Gesichtserkennung - Auflösung ist alles

3.9 Branderkennung - Kontrast muss sein

3.10 Zählung

4 Praxisbeispiele aus vier Anwendungsbereichen

4.1 Der Bahnhof

4.2 Flughäfen

4.3 Einzelhandel - Retail

4.4 Banken

4.5 Autobahn- und Stadtverkehr

4.6 Grenzen und Hürden

5 Installations- und Planungshilfe

5.1 Technische Vorbemerkungen

**5.2 Historische Betrachtungsweisen –
Zukünftige Herausforderungen**

**5.3 Praktische Installations- und
Planungshilfe**

**5.4 Analysefunktionen: Kamerafunktion
stetig überprüfen**

6 Videoüberwachung und Datenschutz

6.1 Videoüberwachung durch Unternehmen

**6.2 Zulässige Videoüberwachung auf
öffentlich zugänglichen Flächen**

6.3 Pflicht zur Videoüberwachung

**6.4 Nicht öffentlich zugängliche Bereiche
und Überwachung am Arbeitsplatz**

**6.5 Beweisverwertungsverbot bei
Regelverstoß?**

6.6 Videoüberwachung durch den Staat

**6.7 Ermächtigungsgrundlagen in den
Polizeigesetzen**

**6.8 Videoüberwachung auf öffentlichen
Plätzen**

6.9 Kfz-Kennzeichen-Scanning

7 Illusionen und Mythen

7.1 Der geheimnisvolle Gang des Menschen

7.2 Bin Laden unter 6 Milliarden Menschen

7.3 Tracken in der Schrägperspektive

7.4 Laufen Bombenleger anders?

7.5 Der Schatten des Hooligan

7.6 Der böse Blick

7.7 Diebe sind schnell

Schlusswort

Sachverzeichnis

Vorwort

Die kleinen und großen Sicherheitszentralen dieser Welt sind verbunden mit Hunderten oder gar Tausenden von Kameras. Diese werden meist von Dutzenden von Monitoren oder durch ganze Videowände visualisiert. Davor sitzen Beobachter, die die Sicherheit von Flughäfen, Bahnhöfen oder öffentlichen Gebäuden und Plätzen gewährleisten sollen. Wissenschaftliche Untersuchungen haben die physischen und psychischen Grenzen der Aufmerksamkeit analysiert und gezeigt, dass der Mensch nach 15 Minuten nur noch 40 bis 50% des zu beobachtenden Geschehens wahrnimmt. Die Frage ist sicherlich gerechtfertigt, wie Sicherheit über Video überhaupt zu gewährleisten ist: Nach 20 bis 30 Minuten ist es nur noch Glücksache, ob ein Beobachter von Sicherheitsmonitoren eine außergewöhnliche oder gar bedrohliche Situation erkennt und auf sie reagiert.

Das Zauberwort ist "Intelligente Videoanalyse". Nur diese kann Tausende von Videokanälen in Echtzeit "sicher" und dauerhaft überwachen! Der Mensch bleibt hierbei freilich auch weiterhin eine sehr wichtige Instanz: Er kann sich dabei allerdings auf seine Stärken konzentrieren: Er prüft, entscheidet und koordiniert.

Hat man sich entschieden, ein Intelligentes Analysesystem einzusetzen, ist es von elementarer Wichtigkeit, für die benötigten Anforderungen das richtige System auszuwählen! Der Markt der Videoüberwachungsanlagen (englisch: Closed Circuit Television, CCTV) ist sehr innovativ. Die rasante Entwicklung von Soft- und Hardware ist aber auch irritierend und verursacht eine massive Verwirrung bei Planern, Errichtern und vor allem beim Endanwender, der es kaum noch schafft, auf dem neusten Stand der Technik zu sein. Und das ist ihm auch nicht zu verdenken, da es

mangels langjähriger Erfahrungen keinerlei Standards oder Orientierungsmaßstäbe gibt.

Innovationen kommen von allen Seiten: von etablierten CCTV-Marktteilnehmern ebenso wie von Technologiequereinsteigern, aus der Wissenschaft und der IT-Welt. Sie alle werfen uns die verschiedensten Fachbegriffe um die Ohren. Dass man hier als Nichtprogrammierer kaum noch folgen kann, ist verständlich.

Da wir alle nur Menschen sind, brauchen wir Zeit, um uns in dieser neuen Softwareanalyse-Welt einzufinden. Schon die Flut von Namenskreationen wie "Intelligent IP", "Intelligent Video", IVS-, IVA-, CAS- und VCA-System verwirrt. Dabei meinen sie alle dasselbe: "Intelligente Videoanalyse". Selbst so klare und schon oft gehörte Begriffe wie "Tracking" bergen bei näherem Hinsehen mehr Unklarheiten als man meinen könnte: Sie zeigen sich bereits, wenn man einmal detaillierter nachfragt, was genau eigentlich getrackt werden kann. Fragt man weiter, welchen Nutzen dieses Tracking in der täglichen Arbeit bringt, werden die Antworten noch spärlicher.

Um eben diese Feinheiten und Fachtermini, um die Unterschiede bezüglich der Qualitäts- und Leistungsmerkmale von Analysesoftware, geht es in diesem Buch, aber genauso um die Grenzen, die Mythen und Sagen der Intelligenten Videoanalyse, die mindestens genauso wichtig sind. Die Materie lässt sich für den Nichtspezialisten kaum auf den ersten Blick durchschauen. Wir vermitteln dieses Wissen mit Theorie, aber in erster Linie anhand von Beispielen aus der Praxis, um Ihnen das nötige Grundwissen mit auf den Weg zu geben. So können Sie sich auf solider Grundlage für Ihr passendes Intelligentes Analysesystem entscheiden. Dazu bringen wir Ihnen auch die Grundregeln der Planung Intelligenter Analysesoftware nahe, geben Ihnen Kriterien zur Auswahl und Positionierung von Kameras an die Hand und vieles mehr.

Darüber hinaus möchten wir Ihnen zwei für Ihre Entscheidungen wichtige Trends aufzeigen, die den Markt aus unserer Sicht in Zukunft prägen werden. Der erste Trend betrifft die hohe Wirtschaftlichkeit der Intelligenten Videoanalyse: Sie hat nicht nur Auswirkungen auf die Entwicklung der traditionellen Sicherheitsmärkte. Schritt für Schritt werden auch andere Märkte — z. B. Einzelhandel (Retail), Banking, Telekommunikation, Logistik und Verkehr, Medizin, Marketing etc. — die Intelligente Videoanalyse für sich entdecken und sie nicht mehr als Kostenfaktor begreifen, sondern als profitablen Geschäftsbereich unter dem Namen "Business Intelligence" entwickeln. Hierzu können Sie in den Anwendungsbeispielen im Kapitel 4 mehr erfahren.

Der zweite Trend betrifft den Wandel in der Entscheidungsebene von Unternehmen. Der Sicherheitsmanager, der heute über das zu integrierende Sicherheitssystem entscheidet, wird mehr und mehr Kompetenzen und Entscheidungen dem IT-Manager überlassen und lediglich als Berater mitwirken. Das liegt daran, dass im Zuge des Technologiewandels die Kameras und Sicherheitssysteme in die Unternehmensnetze eingebunden werden und somit in das Hoheitsgebiet der IT-Welt wandern. Dies birgt neue Chancen und Potentiale von erheblichem Ausmaß sowohl für die klassischen Sicherheitsunternehmen als auch für den IT-Spezialisten.

Dieses Nachschlagewerk soll dazu dienen, Transparenz zu schaffen, den Umgang mit Intelligenten Videoanalyzesystemen zu erleichtern, die neuen Möglichkeiten, aber auch die vorläufigen Grenzen aufzuweisen und der neuen Generation von Videoanalyse positiv entgegenzusehen. Wir wünschen Ihnen hierbei viel Erfolg.

Wir danken recht herzlich für die Unterstützung

Unseren Familien, Axis Communications, AxxonSoft, Bosch Sicherheitssysteme, Brijot Imaging Systems Inc., Dallmeier Electronics, Matthias Erler, Flir, Fraport AG, Geutebrück, Netavis, Mathias Nolte, Klaus Schweizer, TU Berlin, TU Graz, ARS (Wien), Vis-à-pix, Verkehrszentrale Hessen, Wikipedia, Object Video, Fastcom, Franco Baroni und Thomas Bückmann.

Wiesbaden, Januar 2010 Torsten Anstädt, Ivo Keller und
Harald Lutz

1

Historie und wissenschaftliche Perspektive

Wie alles begann: Die Gratwanderung zwischen Möglichkeiten und Nutzen

Um die Gegenwart zu verstehen, ist es meist interessant und auch hilfreich, auf den Ursprung zurückzublicken. Vorab zu bemerken ist, dass es die Entwicklung und die Akzeptanz der Intelligenten Videoanalyse im Gegensatz zu anderen wissenschaftlichen Zweigen immer etwas schwerer hatten. Dies liegt unter anderem an der verbreiteten Angst davor, dass Maschinen (KI-Rechner, KI: künstliche Intelligenz) irgendwann so intelligent wie Menschen sein könnten. Andererseits wurde immer wieder an der Leistungsfähigkeit dieser Technologie gezweifelt zunächst seitens der Wissenschaft selbst, später auch von der Industrie. Beides basiert allerdings in erster Linie auf Unwissenheit!

Seit mehr als 50 Jahren ist die KI-Technologie in vielen Industriezweigen etabliert und nicht mehr wegzudenken, so zum Beispiel in der LCD- oder TFT-Produktion. Dort würde der Wegfall Intelligenter Analyse eine wirtschaftliche Katastrophe bedeuten. Das Gleiche wird man in wenigen Jahren auch von der Sicherheits- und Marketing-Branche behaupten können.

Es begann alles mit Alan Mathison Turing, der 1912 in London geboren wurde und ein britischer Logiker, Mathematiker, Kryptoanalytiker sowie Grundsteinleger der künstlichen Intelligenz war. Turing gilt heute auch als einer

der einflussreichsten Theoretiker der frühen Computerentwicklung und Informatik. Das von ihm entwickelte "Berechenbarkeitsmodell der Turing-Maschine" bildet eines der Fundamente der theoretischen Informatik. Während des Zweiten Weltkrieges war er maßgeblich an der Entzifferung deutscher Funksprüche beteiligt, die mit der Chiffriermaschine "Enigma" verschlüsselt worden waren. Der Großteil seiner Arbeiten blieb nach Kriegsende jedoch unter Verschluss. Er entwickelte 1953 eines der ersten Schachprogramme, dessen Berechnungen er selbst durchführte und zwar wegen mangelnder Hardware-Kapazitäten. Dies ist noch heute ein bekanntes Problem für KI-Entwickler. Nach ihm benannt sind der Turing-Preis, die bedeutendste Auszeichnung in der Informatik, sowie der Turing-Test zum Nachweis künstlicher Intelligenz.

1943 veröffentlichten Warren McCulloch und Walter Pitts im Bulletin of Mathematical Biophysics ihren Aufsatz. "A logical calculus of the ideas immanent in nervous activity". In ihm entwarfen sie die Idee eines Rechenwerkes auf der Basis von Neuronen und Feedback-Schleifen. Es sollte wie eine Turing-Maschine arbeiten und wurde von Alan Turing erstmals im Jahre 1936 beschrieben. 1951 bauten Marvin Minsky und Dean Edmonds den SNARC (Stochastic Neural Analog Reinforcement Calculator), einen neuronalen Netzcomputer, der das Verhalten einer Maus in einem Labyrinth simulierte. Etwas ähnliches versuchte Claude "Entropy" Shannon 1952 mit seiner Maschinenmaus Theseus zu programmieren.

Der Begriff künstliche Intelligenz (englisch: Artificial Intelligence, AI) tauchte erstmals 1955 auf. Geprägt hat ihn John McCarthy in einem Förderantrag an die Rockefeller Foundation für einen 2-monatigen Workshop zu diesem Thema. Er organisierte am 13. Juli 1956 die berühmte Dartmouth Conference am Dartmouth College, die erste Konferenz überhaupt, die sich dem Thema künstliche

Intelligenz widmete. Was auf der Dartmouth-Konferenz entdeckt wurde, war die schlichte Tatsache, dass Computer mehr können als nur komplizierte Ballistik-Formeln zu berechnen. Es war die Entdeckung, dass Computer auch mit Symbolen und Begriffen umgehen können. Das Logical-Theorist-Programm, entwickelt vom späteren Nobelpreisträger Herbert Simon und Allen Newell, war erstmals dazu in der Lage, eine Menge von logischen Theoremen zu beweisen. Konkret führte der Logical Theorist den Beweis von 38 Theoremen aus der *Principia Mathematica* von Bertrand Russell und Alfred North Whitehead. Dieses Ergebnis war ein Meilenstein der künstlichen Intelligenz, da gezeigt wurde, dass Programme zu Aktionen fähig sind, für die ein Mensch Intelligenz braucht.

Herbert Simon prognostizierte 1957 unter anderem, dass innerhalb der nächsten zehn Jahre ein Computer Schachweltmeister werden sowie einen wichtigen mathematischen Satz entdecken und beweisen würde. Beides waren Prognosen, die nicht eintrafen und die Simon 1990, diesmal allerdings ohne Zeitangabe, wiederholte. Immerhin gelang es 1997 dem von IBM entwickelten System "Deep Blue", den Schach-Weltmeister Garry Kasparov in sechs Partien zu schlagen. Unter den zehn ersten Teilnehmern an der Dartmouth-Konferenz, die die KI-Forschung in den nächsten 20 Jahren prägen sollten, gehörten unter anderem Herbert Simon, Marvin Minsky und John McCarthy. Sie gründeten das AI-Lab am Massachusetts Institute of Technology (MIT), aus dem in den folgenden Jahren und Jahrzehnten eine ganze Reihe bahnbrechender Entwicklungen hervorging.

Ray Solomoff entwickelte die algorithmische Informationstheorie. Oliver Selfridge legte mit seinem Pandemonium-Modell zur automatischen Mustererkennung die Grundlagen zur aspektorientierten Programmierung.

Trenchard More entwickelte eine Array-Theorie für die Programmiersprache APL und Arthur Samuel wurde mit seinem selbstlernenden Dame-Spielprogramm bekannt. Sie alle wurden mit ihren Forschungen, Konzepten und Entwicklungen zu Pionieren im Bereich der künstlichen Intelligenz.

2

Algorithmen der intelligenten Videoanalyse

Das menschliche Sehen ist ein hoch komplexer Vorgang, der bisher nur teilweise verstanden wurde. Seiner Erforschung widmen sich die Fachdisziplinen Wahrnehmungspsychologie, die stark technisch ausgerichtete Computer Vision hier geht es um das Sehvermögen von Computern und schließlich die Gebiete des Verstehens: die Semantik und die Kognitionswissenschaften. All diese Fachrichtungen nehmen wiederum Anleihen aus den diversen Disziplinen der Mathematik und Physik, aus der Feldtheorie, der Thermodynamik, der Werkstofftechnik und vielen mehr. Ein Video-Algorithmus setzt sich daher aus zig einzelnen Verfahren zusammen, von denen einige wesentliche Bausteine hier vorgestellt werden.

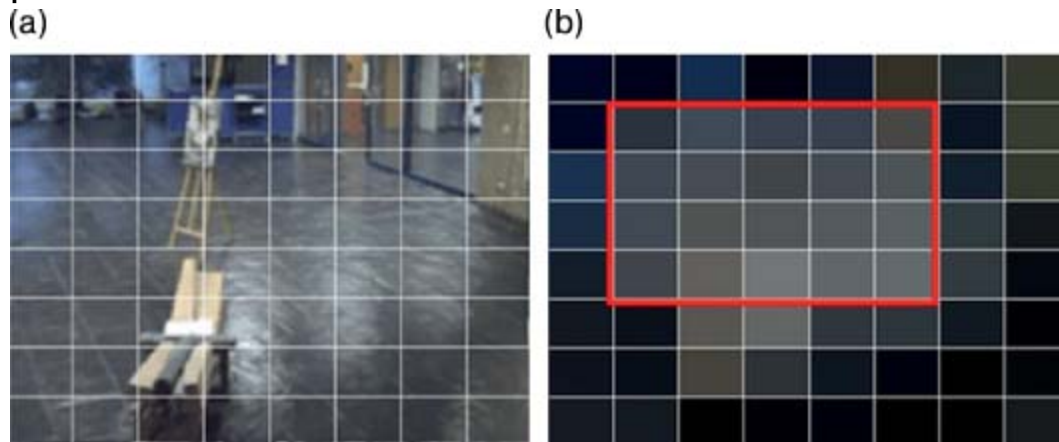
2.1 Klassische Bewegungserkennung Motion Detection

Das Rezept: Unterteile das Bild in Kacheln, beobachte die Farbwerte und melde Änderungen

Im Perimeterschutz (Umfeldschutz eines Gebäudes oder einer Anlage) findet sich noch häufig das älteste Analyseverfahren: Motion Detection. Hierbei wird die Szene in einzelne Kacheln unterteilt ([Abb. 2.1a und b](#)). In der Anlernphase wird deren Mittelwert und typischer

Rauschpegel beobachtet. Anschließend werden die einzelnen Kacheln einer Regelpolitik unterworfen. So lassen sich Kacheln uhrzeitabhängig scharf schalten. ändert sich der mittlere Farbwert oder der Rauschpegel, wird Alarm ausgelöst.

Abb. 2.1 a: Szene in Kacheln unterteilt, b: Analyse der mittleren Farbwerte in scharf geschalteten Kacheln. Vis-à-pix, Fraunhofer HHI



Heutige Motion-Detection-Verfahren erlauben das Zeichnen beliebiger sensitiver Flächen per Mausklick. Sie lassen sich untereinander völlig frei kombinieren. So kann beispielsweise ein Versorgungsweg tagsüber unbeobachtet bleiben, mit Einschalten der Beleuchtung aber scharf geschaltet werden. Ebenso lassen sich Objektgrößen definieren, wodurch sich Fahrzeuge von Personen oder kleinen Tieren unterscheiden lassen.

Beim so genannten Motion-Tracking-Verfahren werden die Farbwerte von Kachel zu Kachel verfolgt ([Abb. 2.2](#)). Damit lassen sich erlaubte und verbotene Richtungen definieren.

Abb. 2.2 Farbänderung beim Motion Tracking. Vis-à-pix, Fraunhofer HHI



Beispielsweise sollen sich im Empfangsbereich morgens alle Personen frei bewegen dürfen, vormittags dagegen müssen sich die Besucher zunächst bei der Empfangsdame melden, abends dürfen die Personen den Raum nur verlassen.

Für einfache Innenbereichs-Szenarien reichen diese Verfahren meist aus. Unter konstanter Beleuchtung und bei wenig Schatten oder Spiegelungen lassen sich große von kleinen Objekten unterscheiden. Im Außenbereich benötigt man zunächst niedrige, windfeste Bepflanzung sowie eine stabile Wetterlage oder dauerhafte nächtliche Beleuchtung. Die Verfahren erlernen die Bewegungen von Büschen, Bäumen, Rasen, Schneeflocken, Schattenschlag oder Regen. Ist dies geschehen, können Wetteränderungen beispielsweise keinen Alarm mehr hervorrufen.

Die Kameramontage ist bei der Anwendung solcher Verfahren von essentieller Bedeutung. Schneeflocken oder Regentropfen, die frontal auf die Linse treffen oder großflächig direkt beleuchtete Schneeflocken könnten zu Auslösungen führen. Bei konfigurationsfreien Algorithmen muss man sich auch darüber im Klaren sein, dass Vögel oder Wild sich ebenfalls vom gelernten Hintergrund deutlich unterscheiden und daher einen Alarm auslösen können.

Motion-Detection-Verfahren ([Abb. 2.3](#)) können bereits einfache Verhaltensmuster erkennen und einen Alarm auslösen, wenn etwa ein Fahrzeug länger als drei Minuten an einer vorgegebenen Stelle parkt oder wenn ein Gegenstand zurückgelassen wird (statische Geschehnisse). Diese Verfahren sind vergleichsweise einfach, robust und benötigen nur geringe Rechenleistung. Für Aufgaben über

den einfachen Perimeterschutz hinaus sind sie dagegen nicht geeignet.

Abb. 2.3 Erweiterte Motion Detection.



2.2 Personenmodelle

Das Rezept: Personenmerkmale aus allen Lebenslagen

Will der Rechner Personen und ihr Verhalten analysieren, benötigt er Personenmodelle. Im einfachsten Fall handelt es sich um Schablonen, die in die Szene eingepasst werden. Eine solche Schablone schiebt man virtuell solange über die Szene, bis man Übereinstimmungen feststellt. Leistungsfähiger ist jedoch nicht der direkte Vergleich im Bild, sondern ein Vergleich der Merkmale. Hierbei besteht eine Person aus Kanten und Formen ("Merkmalen"). Die nachfolgenden Bilder zeigen einige der Personenmodelle, die derzeit durch die internationale Forschergemeinde zirkulieren ([Abb. 2.4](#), [2.5](#) und [2.6](#)).

Abb. 2.4 Ermittlung des Verhältnisses Höhe zu Breite. D. Damen, D. Hoog "Detecting carried objects in short video sequences", ECCV, Part III, S. 156-167, 2008

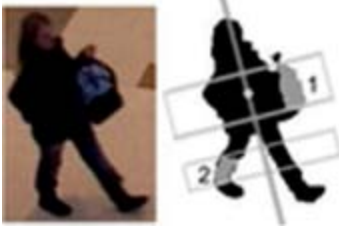


Abb. 2.5 Formen- und Kantenmerkmale der Person im mittleren Bild. N. Dalal, B. Triggs “Histograms of oriented gradients for human detection”, CVPR, S. 886–893, 2005



Abb. 2.6 Skelettmodell in ausgewählten Szenarien. S. Gammeter, A. Ess, T. Jäggli, K. Schindler, B. Leibe, L.V. Gool, “Articulated multi-body tracking under egomotion” ECCV, Part II, S. 816830, 2008



Die oben gezeigten Modelle eignen sich für Detektions- und Tracking-Aufgaben. Hierbei sind Personen nur zu “verfolgen”, was aber bereits bei lockeren Menschenmengen einen hohen Rechenaufwand erfordert.

Die in [Abbildung 2.6](#) gezeigten Skelettmodelle sind bei aller Anschaulichkeit bisher nur für Laborszenarien berechenbar, wobei hier die Echtzeitbedingung weit verfehlt wird. Zwar hofft man, insbesondere durch die Analyse der