

Martin Drahanický, Tomáš Goldmann, Martin Spurný

# Gesichtsdetektion und -erkennung in Videos aus öffentlichen Kamerasystemen

## Inwieweit lassen sich Daten aus öffentlichen Kamerasystemen für polizeiliche Zwecke verwenden?

Können wir das Gesicht des Täters in jeder Situation erkennen? Was benötigen wir dazu? Das Ziel dieses Artikels ist es, die Leser mit den Aufsichtskamerasystemen bekannt zu machen und deren jeweiligen Einsatz in der Praxis sowie die Anwendbarkeit für polizeiliche Zwecke (z. B. Ermittlungen bei Straftaten) zu diskutieren.

### 1 Übersicht über die Kamerasysteme

Dieses Kapitel widmet sich der Übersicht über die gegenwärtigen Kamerasysteme sowie deren Einstellung und Verwendung vor allem für die kriminalpolizeiliche Praxis. Der Umfang der

Nutzung von Aufzeichnungen im Rahmen der Straftatenermittlung wächst enorm, weil auch der Einsatz von Kamerasystemen einen Aufschwung erlebt.

#### 1.1 Übersicht über die Kamerasysteme

Die Sicherheitskamerasysteme können entsprechend der Art der Datenübertragung in analoge und digitale eingeteilt werden.

*Analoge Signalübertragung* – das Signal aus der Kamera kommt in analoger Form mittels eines asymmetrischen Koaxialkabels, unter Verwendung der Fernsehnormen PAL oder NTSC. Die Vorteile der analogen Signalübertragung und der damit verbundenen benutzten Komponenten liegen vor allem in der einfachen Installation. Dennoch überwiegen die Nachteile (unzureichende Bildauflösungen) alle Vorteile, und das auch dann, wenn wir Kameras und Aufzeichnungsgeräte von höchster Qualität einsetzen, wo die horizontale Auflösung auf annähernde DVD-Qualität beschränkt ist. Also eine nachträgliche Signalverarbeitung, die – wie in amerikanischen Krimiserien – zur Unterscheidung von feinen Details führt, ist nicht machbar. Die Bedienungssoftware ermöglicht zwar eine Vergrößerung gespeicherter Videosequenzen; mit der Vergrößerung gehen aber Schärfe und winzige Details verloren, d. h. sie bleiben weiterhin schlecht erkennbar.

*Digitale Signalübertragung* – digitale Systeme kennen keine theoretische Beschränkung der maximalen Bildauflösung von Kamerasystemen. Die maximale Auflösung ist nur von Kameratyp und Datendurchsatz des Netzwerks abhängig. Diese Systeme arbeiten meist mit der klassischen Datenübertragung über das TCP/IP-Protokoll. Der größte Vorteil der digitalen Signalübertragung ist nicht nur die Auflösung (also mehr Bildinformationen), sondern auch die Sicherheit, weil im Gegensatz zu den klassischen analogen CCTV-Systemen die Bedienungs- und Aufnahmegeräte nicht in demselben Haus installiert werden müssen wie die Kameras. Damit wachsen aber auch die Sicherheitsanforderungen an die Datenübertragung, vor allem bei den TCP/IP-Kamerasystemen.



**Martin Drahanický**

arbeitet als Dozent an der Fakultät für Informationstechnologien (CZ) der Technischen Universität in Brunn. Weitere Informationen unter <http://www.fit.vutbr.cz/~drahan>

E-Mail: drahan@fit.vutbr.cz



**Martin Spurný**

arbeitet seit elf Jahren bei der tschechischen Kriminalpolizei als Forensik-Experte für Bild- und Video-Verarbeitung. Zudem ist er an der Brünner Masaryk-Universität als Doktorand eingeschrieben

E-Mail: martin.spurny@pcr.cz



**Tomáš Goldmann**

ist Doktorand an der Fakultät für Informationstechnologien (CZ) der Technischen Universität in Brunn. Sein Hauptinteresse gilt den biometrischen Systemen, der Algorithmenakzeleration und besonders der Gesichtserkennung

E-Mail: igoldmann@fit.vutbr.cz

## 1.2 Nutzbarkeit von Kamerasystemen

In diesem Artikel konzentrieren wir uns auf öffentlich zugängliche Plätze, wo es am häufigsten zu widerrechtlichen Handlungen kommt. Es geht vor allem um die Überwachung von vordefinierten Sicherheitszonen in Gebieten mit erhöhtem Vorkommen von Gesetzesverstößen; es geht aber auch um Prävention. Übrigens haben die meisten gegenwärtig installierten Kamerasysteme lediglich präventive Funktion, weil sie nicht über die notwendigen Eigenschaften und Parameter verfügen, die eine einwandfreie Analyse, Verarbeitung und Verwendung der Videoaufnahmen als Beweismittel ermöglichen – auch nicht auf dem Gebiet der nachträglichen Untersuchung im Rahmen der Bildauswertung.

Der wirkliche Nutzen des Kamerasystemprodukts, d. h. der Videoaufnahme als Beweismittel, liegt überwiegend in der Aufnahme und Abspeicherung von Bild- und Toninformationen; diese enthalten Hinweise nicht nur über den Vorgang in der dargestellten Zeit (und Lokalität), sondern können vor allem auch Entstehung, Verlauf und Vollendung der widerrechtlichen Tat klären und damit verbundene spezifische Merkmale, Charakteristiken und Eigenschaften des aufgezeichneten Objekts darstellen.

### 1.3 Einstellung der Kamerasysteme

Ein entscheidender Punkt bei der Verwendbarkeit von Videoaufzeichnungen unter Zuhilfenahme von Sicherheitskamerasystemen ist deren Einstellung, die gerade von der eingesetzten Technologie für die Aufnahme der realen Welt limitiert wird. Die wichtigsten Parameter einer richtigen und nutzbaren Einstellung von Sicherheitskamerasystemen sind *Aufnahmegröße*, *Auflösung*, *Bildfrequenz*, *Kameraposition* und nicht zuletzt die *Lichtbedingungen*. Falls wir bei den meistverbreiteten Systemen bleiben, d. h. bei Systemen mit analoger Signalübertragung, müssen wir leider konstatieren, dass der PAL- (576i) oder der NTSC-Standard (486i) deren maximale Auflösung begrenzt. Diese Standards werden auch heute im digitalen Zeitalter, wo Auflösungen wie FullHD (1080p) und 4K-UltraHD (2160p) verbreitet sind, noch immer als Standardauflösung für Kamerasysteme betrachtet und genutzt [1]. Ebenso hat die Größe der Aufnahme, definiert im Rahmen des internationalen Standards [1] und eingeteilt in *Überwachung*, *Entdeckung*, *Beobachtung*, *Erkundung*, *Identifikation* und *Untersuchung* (siehe Abb. 1.1), in ihrem vollen Umfang nur bei digitalen Systemen mit digitaler Signalübertragung reale Bedeutung für die Objektidentifikation. Bei den analogen Systemen kann unter realen Bedingungen nur mit jenen Aufnahmegrößen gearbeitet werden, die als *Erkundung*, *Identifikation* und *Untersuchung* definiert werden [2].

**Abbildung 1.1 | Vergleich von vier Aufnahmegrößen (v.l.: Identifikation, Erkundung, Entdeckung, Überwachung).**



Nicht zuletzt muss auch die eingestellte minimale Bildfrequenz berücksichtigt werden, die mindestens 15 fps (Bilder pro Sekunde) betragen sollte. Für den Fall einer Detektion des überwachten Ereignisses sollte die Bildfrequenz 25 fps erreichen.

## 1.4 Beispiele aus der polizeilichen Praxis

Im Moment überwiegen leider noch die Sicherheitskamerasysteme mit analoger Datenübertragung. Damit sind auch die meisten Probleme verbunden, da die gespeicherten Daten (in den meisten Fällen handelt es sich um dynamische Bilddaten, die oft fälschlicherweise als Videosequenzen bezeichnet werden) als Beweismittel im Strafverfahren und in der Gerichtsverhandlung nicht verwendbar sind. Dasselbe gilt aber auch für Kamerasysteme mit digitaler Signalübertragung, falls wichtige Eigenschaften oder Einstellungen missachtet werden.

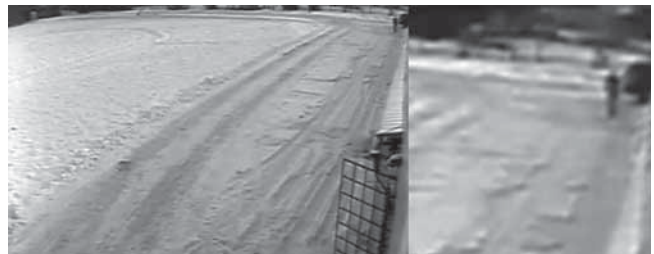
### 1.4.1 Aufnahmegröße

Ein typisches Beispiel einer unpassend gewählten Aufnahmegröße ist der Einsatz von Stadtkamerasystemen, bei denen die Aufnahmegröße in vielen Fällen nur auf die Überwachung eingestellt wird. Durch die unpassend eingestellte maximale Auflösung des Aufnahmegeräts auf die Größe CIF (352 × 288 Pixel) und die standardmäßig völlig ungeeignete Kameraposition entsteht eine Videoaufnahme, die als Beweismittel nicht verwendet werden kann. Beispiele für ungeeignet eingestellte Stadtkamerasysteme sind in Abb. 1.2 gezeigt, wo eine ungeeignete Aufnahmegröße, eine niedrige Auflösung und eine ungeeignete Kameraposition gewählt wurden. Weitere typische Beispiele von ungeeignet gewählten Aufnahmegrößen sind die sog. Straßenkamerasysteme, die an den Fassaden von Wohn- und Nichtwohnhäusern installiert werden (siehe Abb. 1.3).

**Abbildung 1.2 | Stadtkamerasystem mit ungeeigneter Aufnahmegröße, Auflösung und Kameraposition.**



**Abbildung 1.3 | Aufnahme einer an einer Hausfassade installierten Kamera.**



### 1.4.2 Auflösung

Bei der Auflösung kann festgehalten werden, dass dort, wo die Grenzen der analogen Signalübertragung liegen, die Möglichkeiten der digitalen Signalübertragung anfangen. Es ist auch wich-

tig zu bemerken, dass die Verwendbarkeit der Videoaufnahme als Beweismittel nicht nur von der Auflösung der Quelle abhängt, sondern auch von der dazugehörigen Aufnahmegröße. Das Beispiel in Abb. 1.4 zeigt zwei Aufnahmen mit geringer Auflösung, was die Identifizierung von Personen oder Autokennzeichen verhindert.

**Abbildung 1.4 | Eine sehr geringe Auflösung der Videoaufnahme verhindert jede Identifikation.**



### 1.4.3 Kameraposition

Die Kameraposition ist bei beiden Signalübertragungen vom Benutzer frei wählbar, was aber oft bei der Installation und Inbetriebnahme unterschätzt wird. Beispiele für ungeeignet positionierte Kameras sind in Abb. 1.5 zu sehen – in diesem Fall ist die Identifizierung des Objekts auf der Aufnahme kaum möglich.

**Abbildung 1.5 | Ungeeignet positionierte Kameras.**



## 2 Beschreibung der HeadViewer-Applikation

Der Erfolg der Gesichtsdetektion wird von vielen Faktoren beeinflusst, die wir in zwei Gruppen einteilen können. Die erste Gruppe enthält solche Faktoren, die wir hinsichtlich der Gesichtserkennung nicht beeinflussen können. Dazu gehören Aspekte, die durch die Umgebung bedingt sind, sowie die visuelle Zugänglichkeit des Gesichts, das wir detektieren wollen. Die zweite Gruppe enthält beeinflussbare Faktoren, mit denen wir den Erfolg der Gesichtsdetektion verbessern können. Einer dieser Faktoren ist die Kameraposition im Verhältnis zum Gesicht der zu identifizierenden Person.

Moderne Gesichtsdetektoren sind gewöhnlich auf frontale Gesichtsbilder trainiert, was in der Praxis bei einer nicht frontalen Kopfposition gegenüber der Kamera zu einer geringeren Detektionsmöglichkeit führt. Sehr oft sind die CCTV-Kameras mit einer Drehung bzw. Neigung im Verhältnis zu dem überwachten Platz installiert, was die automatische Gesichtsdetektion in den Aufnahmen dieser Kamerasysteme verschlechtert. Falls es zu einer Detektionsverschlechterung kommt, sinkt die Chance, eine Person auf der Aufnahme zu erkennen.

**Abbildung 2.1 | Screenshot der HeadViewer-Applikation.**



Die HeadViewer-Applikation (siehe Abb. 2.1) ist aus dem Bemühen entstanden, die Kameraposition zu optimieren. Mit ihr kann die Kameraposition gegenüber einer Person simuliert werden; anhand der gewonnenen Daten lässt sich dann die Zuverlässigkeit der Gesichtsdetektion bestimmen. Bei der Kamerainstallation ist im Anschluss unbedingt zu überprüfen, ob die Grenzwerte (Winkel) überschritten wurden.

Die Applikation ist auf experimenteller Basis mit Gesichtsdetektoren entstanden, die auf unsere einzigartige Gesichtsdatenbank appliziert wurden. Die Einzigartigkeit der Datenbank liegt darin, dass wir bei jedem Bild eine genaue Position des Gesichts gegenüber die Kamera abgespeichert haben. Diese Datenbank wurde zum Zwecke der Entwicklung dieser Applikation erstellt. Wir verfügen hier über mehr als 120 Bilder von drei unterschiedlichen Personen, von denen jedes mit einem Etalon versehen ist. Für die Gesichtsdetektion haben wir Detektoren gewählt, die Haar-Merkmale (siehe Unterkapitel 3.1) und das Histogramm der orientierten Gradienten (siehe Unterkapitel 3.1) verwenden.

Die Applikation ermöglicht die Einstellung der Kamerahöhe, der Entfernung zwischen Person und Kamera, der Größe der Person, der geschätzten Trajektorie der Bewegung der Person und den Kamerawinkel im Verhältnis zu dem Gesicht.

Der wichtigste Teil der Applikation ist die Anzeige der Gesichtsdetektionszuverlässigkeit. Zur besseren Anschaulichkeit wurden drei Stufen der Gesichtsdetektionszuverlässigkeit definiert. Die erste Stufe repräsentiert das Sensitivitätsintervall zwischen 0.85-1.0, d. h. die Wahrscheinlichkeit der Gesichtsdetektion ist beim Erreichen dieser Stufe sehr hoch. Die nächste Stufe ist für das Sensitivitätsintervall 0.65-0.85 gedacht, wo es aber bei maschineller Gesichtsdetektion zu Ausfällen kommen kann. Falls sich die Einstellungen in der Applikation der dritten Stufe annähern, wäre die Wahrscheinlichkeit der Gesichtsdetektion vollkommen vom Zufall abhängig. Die Auswertung verläuft in der horizontalen und mittleren Ebene; die Applikation kann Azimuts- und Elevationsänderungen simulieren. Als Ergebnis kommt der minimale Wert des Detektionserfolgs aus diesen Ebenen heraus.

Damit der Nutzer bei der Einstellung der Applikation jederzeit eine Übersicht über das Ergebnis hat, enthält die Applikation bei der Einstellung der Höhe und der Trajektorie einen Streifen mit einem Farbgradienten, der die Zuverlässigkeit der Detektion darstellt. Die erste Stufe (die zuverlässigste Detektion) wird mit grüner Farbe, die zweite Stufe mit gelber Farbe und die dritte dann mit brauner Farbe dargestellt.

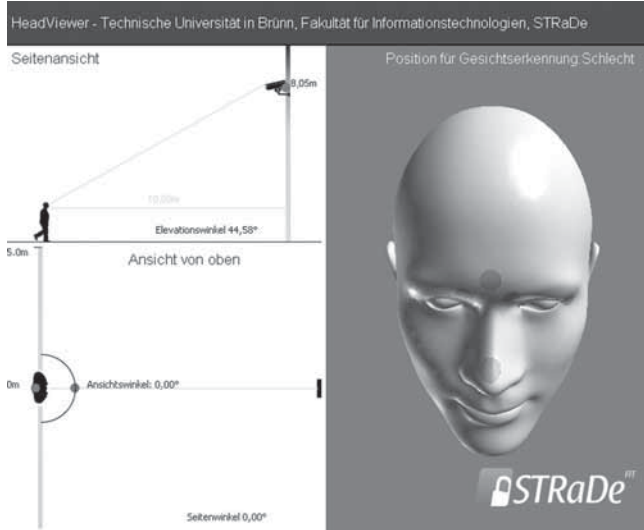
Weil einige Algorithmen für die Gesichtserkennung mit einem Gitter arbeiten, das die nodalen Gesichtspunkte verbindet (siehe Abb. 2.2 und [3]), bietet auch unsere Applikation die Anzeige dieses Gitters an, das durch die Verbindung der grünen Punkte auf dem Gesichtsmodell entsteht.

**Abbildung 2.2 | Beispiel für die nodalen Gitterpunkte auf dem Gesicht aus unterschiedlichen Winkeln**



Die Übersicht über die jeweiligen Kopf- und Kamerapositionen kann in eine Bilddatei exportiert werden. Das Ergebnisbild enthält die Situationsaufzeichnung und die erwartete Zuverlässigkeit der maschinellen Detektion – siehe Abb. 2.3.

**Abbildung 2.3 | Ergebnisbild der HeadViewer-Applikation.**



Die Oberfläche und die Bedienung der Applikation sollen äußerst intuitiv sein, wobei gegenwärtig drei Sprachen unterstützt werden – Deutsch, Englisch und Tschechisch. In naher Zukunft wird die Applikation um weitere wichtige Punkte erweitert, die den Einsatz von CCTV-Systemen effizienter gestalten können. Aktuell arbeiten wir an der Simulation der Aufnahmegröße der Kamera, d. h. die Darstellung des Kopfmodells wird einige Parameter einer konkreten Kamera berücksichtigen, z. B. Kameraauflösung und Verzerrung des Objektivs.

Darüber hinaus möchten wir diese Applikation auf mobile Geräte ausweiten, damit sie vor Ort einsetzbar ist und so die Installation der CCTV-Kamera konkret vereinfacht. Diese Lösung wird die Erstellung einer sog. Heatmap der Wahrscheinlichkeit einer erfolgreichen Gesichtsdetektion in der Aufnahme ermöglichen. Mittels dieser Applikation wird die Verwendbarkeit für eine maschinelle Datenverarbeitung von CCTV-Aufnahmen verbessert.

## 3 Gesichtserkennung mittels U-Rampe und Quadrokopter

Die Verfügbarkeit von Quadroköptern (allgemein Drohnen – die Anzahl der Propeller spielt hier keine Rolle) bringt neue Möglichkeiten in der Personenbeobachtung und -verfolgung. Damit einhergehend entstehen neue Herausforderungen für die Algorithmen (und Schablonen) für die Personenbeobachtung und -erkennung. In den folgenden Teilen werden die Technologien beschrieben, die mit der Aufnahme eines komplexen Gesichtsmodells und mit dem Einsatz von Quadroköptern in Sicherheitssystemen zusammenhängen.

### 3.1 Detektion- und Erkennungsalgorithmen für 2D- und 3D-Gesichter

#### Algorithmen für die Gesichtsdetektion

Zu den meistverwendeten Gesichtsdetektoren in 2D-Bildern gehören zwei, die noch weiter verbessert werden können, nämlich die *Haar-Merkmale* und das *Histogramm der orientierten Gradienten*.

Der Algorithmus, der die Haar-Merkmale verwendet, nennt sich auch Viola-Jones [4] und wurde 2004 entwickelt. Die benutzten sog. Haar-Merkmale stammen aus der Haar-Funktion. Für die Berechnung dieser Merkmale wird das Integralbild herangezogen, das eine schnelle Berechnung dieser Merkmale ermöglicht. Für das Training dieses Detektors wird der Klassifizierungsalgorithmus AdaBoost [5] benutzt. Der Input dieses Klassifizierungsalgorithmus ist die Trainingsmenge, die positive und negative Muster enthält. Aus jedem Muster wird eine Menge von einfachen Merkmalen generiert. Der finale Klassifikator wird aus der Menge der schwachen Klassifikatoren gebildet, die in einer Kaskade angeordnet sind. Der Vorteil des Viola-Jones-Algorithmus liegt in der Unabhängigkeit von Beleuchtung und sinnvoller Objektgröße.

Der zweite Algorithmus, das sog. Histogramm der orientierten Gradienten – HOG (*Histograms of Oriented Gradients*) [6] – wurde zum ersten Mal für die Detektion einer Person anhand der Silhouette verwendet. Falls wir andere Objekte als eine Personensilhouette detektieren wollen, muss das Objekt mithilfe von Kanten gut beschreibbar sein. Das Gesichtsbild erfüllt diese Bedingung problemlos.

Der Algorithmus funktioniert so, dass das Eingangsbild in Grautöne umgewandelt wird, die anschließend in das Bild der ersten Derivation in  $x$ - und  $y$ -Richtungen transformiert werden. Aus diesen Daten können wir die Richtung und Vektorgröße einer konkreten Kante berechnen. Ein wichtiger Bestandteil des Algorithmus ist die Aufteilung des Bildes in Unterbilder. In der ersten Version (Dallal und Triggs) wurde die Aufteilung auf Zellen mit der Größe  $8 \times 8$  verwendet. Aus jeder dieser Zellen wird dann ein Histogramm berechnet, das die Kantenaufteilung im Bild darstellt. Die Erstellung des Histogramms kann auf zwei Arten realisiert werden: (a) einzelne Winkel tragen mit der Steigungsgröße zu einer eigenen Klasse bei; (b) Verteilung der Steigungsgröße zwischen zwei benachbarten Klassen. Falls wir z. B. den Winkel der  $80^\circ$ -Klasse mit dem Mittelpunkt  $70^\circ$  und  $90^\circ$  haben, wird in jede Klasse die Hälfte der Steigungsgröße eingefügt. Falls wir den Winkel  $85^\circ$  haben, wird in diese Klasse ein Viertel des  $70^\circ$ -Anteils und drei Viertel des  $90^\circ$ -Anteils eingefügt. Der Vorteil dieses Ansatzes liegt in der Lösung des Problems, das ent-

stehen würde, wenn der Gradient mit einer markanten Steigung auf der Grenze zweier Klassen läge. Damit wäre die Größe einer Klasse stark beeinflusst, wobei dieser Gradient kaum zu der anderen beitragen würde.

Aufgrund der Begrenzung der Einflüsse der Lumineszenz auf den finalen Deskriptor werden einzelne Histogramme normalisiert. Die Normalisierung wird in den Blöcken durchgeführt, die als eine Verbindung der Histogramme der orientierten Gradienten aus einzelnen Zellen entstehen. Dabei muss erwähnt werden, dass diese Blöcke nicht unbedingt beieinander liegen müssen; sie können sich überlappen oder sogar eine ganz andere Verteilung haben.

#### Algorithmen für die Personenerkennung anhand des Gesichts

Für die 2D-Gesichtserkennung werden Algorithmen verwendet, die auf der Analyse der Hauptkomponenten basieren – der sog. PCA (*Principal Component Analysis*), wo eine Reduktion der Anzahl der Variablen vor der Klassifizierung erfolgt. Einer der Algorithmen, der diese Methode zur Gesichtserkennung benutzt, ist der Algorithmus Eigenface [7]. Dieser arbeitet mit mittelwertgebildeten, monochromatischen Bildern, wobei zu jedem Bild der Eigenface-Vektor gesucht wird, der für die Klassifizierung benutzt wird. Der nächste Algorithmus für die 2D-Gesichtserkennung ist der sog. Fisherface-Algorithmus [8]. Dieser Algorithmus arbeitet mit der Diskriminantenanalyse LDA (*Linear Discriminant Analysis*). Das Hauptziel dieser Methode ist die Aufteilung der Objekte in Klassen auf der Grundlage einer Entscheidungsregel. Das ist der Hauptunterschied zu dem PCA-Algorithmus, der unüberwacht lernt. Der Vorteil des überwachten Lernens liegt darin, dass wir solche Regeln wählen können, die einzelne Unterräume optimal voneinander separieren. Der Nachteil liegt aber darin, dass es zu einer übermäßigen Adaptation des Klassifikators auf eine Person kommen kann.

Für die Gesichtserkennung anhand von räumlichen Daten kann die Analyse der Krümmung einzelner Gesichtsteile Anwendung finden. Diese Methode nutzt die Gauß-Krümmung für die Erkennung, die als Produkt zweier Hauptkrümmungen an einer konkreten Stelle definiert wird. Darüber hinaus wird in diesem Punkt der Mittelwert beider Krümmungen berechnet. Der finale Anteil dieser zwei Werte wird verwendet, wobei somit der charakteristische Wert für die Klassifizierung entsteht [9].

Die weiteren Algorithmen arbeiten direkt mit dem 3D-Modell. Für die direkte Arbeit mit dem 3D-Modell wird der ICP-Algorithmus (*Iterative Closest Point*) [10] verwendet. Hier wird die Gesamtentfernung zwischen beiden Mengen mit Punkten im Raum minimiert.

Es gibt mehrere Algorithmen, die mit dem 3D-Modell arbeiten, darunter auch Modifikationen der Algorithmen Eigenface und HOG für 3D-Modelle.

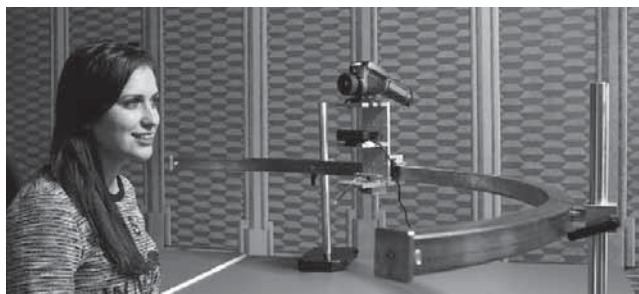
### 3.2 Verwendung der U-Rampe

Mit den Anforderungen an ein genaues 3D-Modell wachsen natürlich auch die Anforderungen an die Datenaufnahme. Das Raummodell des Gesichts können wir mithilfe einer 3D-Kamera bekommen. Der Nachteil liegt aber darin, dass diese Kamera die Strukturdaten in nur einer Ebene aufnehmen kann.

Weil wir ein sehr genaues 3D-Gesichtsmodell schaffen wollen, haben wir ein Gerät – die U-Rampe (siehe Abb. 3.1) – entwickelt. Dieses Gerät besteht aus einer zentralsymmetrischen Bahn,

auf der sich ein Wägelchen bewegt. Dieses Wägelchen wird von einem Schrittmotor betrieben und von einem Computer gesteuert, in den auch die Daten aus den angeschlossenen Kamerasystemen übertragen werden. Dieses Wägelchen kann unterschiedliche Kamerasysteme (RGB-Kamera, Thermokamera oder 3D-Kamera) tragen. Zu jedem Zeitpunkt wissen wir exakt, in welcher Position dieses Wägelchen steht, was uns in die Lage versetzt, ein genaues 3D-Modell aus den einzelnen Daten zusammenzusetzen.

Abbildung 3.1 | U-Rampe.



Bei den ersten Versuchen mit der U-Rampe wurden die Thermokamera FLIR und die 3D-Kamera Intel RealSense (3D-Modell und Texturinformation) eingesetzt.

Ein interessantes Gebiet für die Verwendung der dreidimensionalen Daten ist die Personendetektion und -erkennung. Dabei handelt es sich um 2D-Bilder, die aus unterschiedlichen Winkeln zum Kopf aufgenommen wurden. Eine solche Bilddatenmenge kann für die Algorithmen für die Gesichtserkennung mit gedrehtem oder geneigtem Kopf im Bild gewinnbringend eingesetzt werden. Zusätzlich lässt sich aus diesen Daten eine Schablone erstellen, die zu weiteren Personenidentifikationszwecken verwendet werden kann.

In naher Zukunft kann diese Technologie einen größeren Beitrag für die kriminalistisch-technische Praxis haben, wo Identifikationsangaben des Täters registriert werden können oder sich diese Technologie für die Bevölkerungsevidenz einsetzen lässt. Die Person, die mithilfe dieses Geräts aufgenommen wird, muss nicht aktiv kooperieren – es genügt völlig, dass diese Person für nur wenige Sekunden den Kopf aufrecht hält. Die gewonnenen Daten können für die Personenerkennung in CCTV-Aufnahmen verwendet werden, z. B. bei der Ermittlung einer Straftat oder bei der Beobachtung von unerwünschten Personen in überwachten Gebieten (Fußballspiele, Demonstrationen usw.).

Ein Vorteil dieses Geräts liegt auch in seinem günstigen Anschaffungspreis, weil es ohne teure Komponenten auskommt.

### 3.3 Verwendung von Quadroptern

In gewöhnlichen CCTV-Systemen wird davon ausgegangen, dass jede Kamera eine statische Position hat. Interessant ist die Verarbeitung von CCTV-Aufnahmen von Kameras ohne feste Position. Hierbei handelt es sich um Kameras, welche die Position verändern können, d. h. die Aufnahmen haben variierende Hintergründe, wo eigentlich jene Methoden, die eine Hintergrundsubtraktion [11] für die Extraktion von Objekten verwenden, nicht einsetzbar sind. Als Beispiel für eine nicht statische Kamera kann eine Kamera auf einem Quadroptern dienen.

Der Quadrokooper bietet viele Möglichkeiten, die Arbeiten der Sicherheitsdienste zu verbessern und zu effektivieren. Mit dem Quadrokooper lassen sich interessante Daten gewinnen – in diesem Fall untersucht der Operator fortlaufend einzelne Objekte in dem überwachten Gebiet, und diese Daten werden dann von dem Quadrokooper in das Verarbeitungszentrum gesandt. Bei dieser Verwendung handelt der Quadrokooper nicht autonom.

Wenn wir hingegen eine autonome Überwachung von einem Gebiet oder einer Person benötigen, kann der Quadrokooper autonom handeln. Als Beispiel dafür kann die Suche nach einer Person mittels Thermokamera dienen. Hier wird der Quadrokooper nach einem lebenden Objekt (einer Person) suchen, wobei die Nachricht, falls die Identifikation eines solchen Objekts erfolgreich ist, den Operator erreichen muss, damit er diese auswerten kann. Ein weiteres Beispiel kann die Suche nach einer konkreten Person in einer Gruppe sein – in diesem Fall werden alle Gesichter im Videosignal automatisch detektiert und erkannt. Der Operator bekommt dann die Information, ob sich diese Person in der Gruppe befindet oder nicht.

Ein wichtiger Aspekt ist die Einhaltung der Gesetze und gültigen Richtlinien in den jeweiligen Ländern, die den Betrieb von unbemannten Luftfahrzeugen regulieren – vor allem in einem autonomen Modus ist der Betrieb oft unmöglich oder zumindest sehr kompliziert und begrenzt.

Es lassen sich auch weitere Applikationen mit der Verwendung von Quadrokoopern denken. Jetzt kommen wir aber zu den Erkenntnissen, die mit der maschinellen Beobachtung einer Person mithilfe von Quadrokoopern verbunden sind, bei der auch eine autonome Überwachung einsetzbar wäre.

### Verfolgung einer Person anhand des Gesichts

Falls wir Personen mittels einer nicht statischen Kamera überwachen und verfolgen möchten, müssen aufgrund der hohen Variabilität der Lokalisierung der Quadrokooper ganz andere Verfahren als bei den CCTV-Kameras eingesetzt werden.

In einigen Fällen kann die Person anhand der Silhouette verfolgt werden. Das funktioniert aber nur dann, wenn nicht mehrere Personen im Bild auftauchen, sonst kommt es zu Ausfällen. Deswegen ist es besser, die Personen anhand des Gesichts zu erkennen und zu verfolgen, weil das Gesicht eine hohe Variabilität zwischen den Klassen aufweist.

Bei einer automatischen Personenverfolgung anhand des Gesichts muss der Quadrokooper in einer korrekten Position zum Kopf gehalten werden. Wie aus der HeadViewer-Applikation ersichtlich wurde, ist die Zuverlässigkeit der Gesichtsdetektion stark von der Position der Kamera gegenüber dem Kopf abhängig. Bei der Steuerung der Quadrokooper müssen die Intervalle für Azimut und Elevation so gehalten werden, dass die Detektionszuverlässigkeit auf höchstem Niveau gehalten wird.

Falls auf dem Quadrokooper eine Kamera ohne Zoomfunktion installiert wird, muss auch die Entfernung des Quadrokoopers von dem Kopf in einem konkreten und engen Bereich gehalten werden. Diese Entfernung hängt von dem verwendeten Objektiv ab. Eine Lösung hierfür bieten sog. PTZ-Kamerasysteme (*Pan Tilt Zoom*), die sich für die automatische Überwachung und Verfolgung eignen.

Weil der Quadrokooper ein fliegendes Propellergerät ist, kommt es während des Fluges zu Vibrationen. Deswegen ist die Bildstabilisierung im Video besonders wichtig. Die Bildstabilisierung kann entweder mechanisch (optische Stabilisierung) oder

digital (über die Software) umgesetzt werden. Die Bedeutung der Stabilisierung steigt bei einer niedrigen Bildfrequenz. Bei diesen Kameras kommt es durch Vibrationen zur Verzerrung des Bildes, d. h. zu einer Senkung der Bilddatenqualität. Die Qualität der Bilddaten beeinflussen natürlich ebenso Umgebung und Beleuchtungsstärke.

In den vergangenen Jahren haben wir einige Tests zur Gesichtsdetektion und -erkennung mittels Quadrokooper (Ar. Drone 2.0) durchgeführt. Obwohl es sich um einen billigen Quadrokooper handelte, war dieser für unsere ersten Tests völlig ausreichend. Für die Videoaufnahme haben wir eine eingebaute Kamera und eine GoPro3-Hero-Kamera verwendet.

Das Ziel des Experiments war der Beweis der Zuverlässigkeit der Detektion und Erkennung einer konkreten Person. Weil diese Kameras keine Zoomfunktion anbieten, mussten wir den Quadrokooper in einer eng begrenzten Entfernung von der Person halten. In der folgenden Tabelle lässt sich sehen, wie sich die Zuverlässigkeit der Detektion im Verhältnis zur Entfernung (rechteckige Projektion) und Elevation ändert. Für die Detektion wurden die Algorithmen Viola-Jones und LBP (*Local Binary Pattern*) [12][13] eingesetzt.

**Tabelle 1 |**

Merkmale + Kamera	Entfernung zur Person [m]	Elevation [°]	Zuverlässigkeit der Detektion [%]
HAAR GoPro	4.00	35	95
LBP GoPro	3.50	35	90
HAAR Integr.	3.75	30	85
LBP Integr.	3.50	30	80

Die Gesichtsdetektion war bei sonnigem Wetter um etwa 30% erfolgreicher als bei bewölktem Wetter, wo die Lichtintensität der Sonne relativ niedrig war.

Für die Gesichtserkennung wurden die Algorithmen Eigenface, Fisherface und LBP [12] verwendet. Generell können wir sagen, dass, falls wir das Gesicht detektieren konnten, auch die Erkennung problemlos verlief. Falls die Person ein Tuch oder andere Kleidung trug, die direkt das Gesicht betraf, kam es zu einer Verschlechterung der Erkennungszuverlässigkeit [13], was auch theoretisch erwartet worden war.

## 4 Zusammenfassung

Wie wir schon am Anfang des Artikels bemerkt haben, gibt es eine große Zahl von Problemen, die mit den in der Praxis eingesetzten Kamerasystemen verbunden sind. Wir möchten die Modelle verwenden, die mittels der U-Rampe für die Detektion und Erkennung von Gesichtern in den Aufnahmen aus CCTV-Kamerasystemen gewonnen werden, wobei die U-Rampe auch in der polizeilichen Praxis eingesetzt wird. Auf dem Gebiet der Quadrokooper planen wir die Erweiterung von Detektoren und Erkennungsalgorithmen für die Personenidentifizierung unter Zuhilfenahme von auf einem Quadrokooper installierten RGB- und Thermokameras. Im Rahmen der Forschung und Entwicklung möchten wir die Algorithmen für die Detektion, Erkennung und Verfolgung von Personen verbessern. Den größten Ertrag sehen wir in der Reduzierung des Einflusses von Kopfdrehungen auf

die Zuverlässigkeit der Gesichtsdetektion und -erkennung, darüber hinaus in der Verwendung von Thermokameras für die Personenverfolgung bei verschlechterten Lichtbedingungen. Unsere Ergebnisse bringen hoffentlich auch den Sicherheits- und Rettungsdiensten einen Gewinn.

## Danksagung

Dieser Artikel ist im Rahmen der folgenden Projekte entstanden: „IT4Innovations Excellence in Science“ Ministerium für Schulwesen der Tschechischen Republik – LQ1602 (National Programme of Sustainability (NPU II)) (CZ); „Tools and methods for video and image processing to improve effectivity of rescue and security services operations“ Ministerium des Inneren der Tschechischen Republik mit der Bezeichnung VI20172020068 (VZ) sowie „Sichere und zuverlässige Computersysteme“ mit der Bezeichnung FIT-S-17-4014 (CZ).

## Literatur

- [1] IEC/TC 79. IEC 62676-4:2014. *Video surveillance systems for use in security applications – Part 4: Application guidelines*. ISBN 978-2-8322-1504-3.
- [2] *Australia and New Zealand Police Recommendations for CCTV Systems*. Australia New Zealand Policing Advisory Agency. 2014, National Institute of Forensic Science Australia, S. 2.
- [3] Vacca J. R.: *Biometric Technologies and Verification Systems*. Butterworth-Heinemann, 2007.
- [4] Viola P., Jones M. J.: *Robust Real-time Face Detection*. International Journal of Computer Vision, Vol. 57, No. 2, 2004, S. 137–154.
- [5] Šochman J., Matas J. *AdaBoost and Face Detection*. Czech Technical University, CZ, 2003.
- [6] Dalal N., Triggs B. *Histograms of Oriented Gradients for Human Detection*. In: 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, IEEE, 2005, S. 886–893.
- [7] Zhang J., Yan Y., Lades M. *Face Recognition: Eigenface, Elastic Matching, and Neural Nets*. In: Proceedings of the IEEE, 1997, 85.9: 1423-1435.
- [8] Belhumeur P.N., Hespanha J.P., Kriegman D.J. *Eigenfaces vs. Fisherfaces: Recognition Using Class Specific Linear Projection*. IEEE Transactions on pattern analysis and machine intelligence, 1997, 19.7: 711-720.
- [9] Colombo A., Cusano C., Schettini R. *3D Face Detection Using Curvature Analysis*. Pattern recognition, 2006, 39.3: 444-455.
- [10] Cook J. et al. *Face Recognition from 3D Data Using Iterative Closest Point Algorithm and Gaussian Mixture Models*. In: Proceedings of 3D Data Processing, Visualization and Transmission, 2004, IEEE. S. 502-509.
- [11] Piccardi M. *Background Subtraction Techniques: A Review*. In: IEEE International Conference on Systems, man and cybernetics, IEEE, 2004, S. 3099–3104.
- [12] Ahonen T., Hadid A., Pietikäinen M. *Face recognition with Local Binary Patterns*. In: European Conference on Computer Vision. Springer Berlin Heidelberg, 2004, S. 469-481.
- [13] Mikundová L. *Biometrické rozpoznávání 2D obličejů z kamerového systému umístěného na kvadroptéře (Biometrische 2D-Gesichtserkennung aus einem auf einem Quadroptopter montierten Kamerasystem)*, Bachelorarbeit, Brno, FIT TU Brno, 2015.

 Springer Vieweg

springer-vieweg.de

# Erstes Fachbuch des CEP auf Deutsch in dieser Tiefe



U. Hedtstück  
**Complex Event Processing**  
 Verarbeitung von Ereignismustern  
 in Datenströmen  
 2017. X, 148 S. 123 Abb. Brosch.  
 € (D) 19,99 | € (A) 20,55 | \*sFr 21,00  
 ISBN 978-3-662-53450-2  
 € 14,99 | \*sFr 16,50  
 ISBN 978-3-662-53451-9 (eBook)

- Macht Algorithmen der vernetzten Welt verständlich. Vermittlung des theoretischen Hintergrunds und Einbettung in die Wissenschaft
- Enthält einen Anhang mit Prädikatenlogik 1. Stufe als Grundlage für regelbasierte Algorithmen

Eine wichtige Aufgabe für die IT der vernetzten Welt ist die maschinelle Auswertung und Verarbeitung von Informationen, die für eine Anwendung relevant sind und übers Netz verschickt werden.

€ (D) sind gebundene Ladenpreise in Deutschland und enthalten 7 % für Printprodukte bzw. 19 % MwSt. für elektronische Produkte. € (A) sind gebundene Ladenpreise in Österreich und enthalten 10 % für Printprodukte bzw. 20 % MwSt. für elektronische Produkte. Die mit \* gekennzeichneten Preise sind unverbindliche Preisempfehlungen und enthalten die landesübliche MwSt. Preisänderungen und Irrtümer vorbehalten.

Jetzt bestellen auf [springer.com/Angebot1](http://springer.com/Angebot1) oder in Ihrer Buchhandlung

Part of **SPRINGER NATURE**