



Gaze Control Strategien zur Überwachung dynamischer Szenen unter Verwendung eines Master-Slave Kamera-Systems

**Studienarbeit
von
Eren Müslüm Dursun**

15. September 2013

Betreuer: Prof. Dr.-Ing. Rainer Stiefelhagen¹

Betreuer: Dr. rer. nat. Wolfgang Hübner

Betreuer: David Münch

Fraunhofer Institut für Optronik, Systemtechnik und Bildauswertung
Gutleuthausstr. 1, 76275 Ettlingen

¹Computer Vision for Human-Computer Interaction Lab - Karlsruhe Institute of Technology

Hiermit erkläre ich, dass ich die vorliegende Arbeit selbstständig angefertigt und keine anderen, als die angegebenen Quellen und Hilfsmittel verwendet habe.

Ettlingen, 15. September 2013

Eren Müslüm Dursun

Kurzfassung

Das Hauptanliegen der Videoanalyse besteht darin, semantisch zu interpretieren, was, wer, wo, wann, wie und warum sich etwas in der beobachteten Szene abspielt. Wenn man dies auf Personen und Objekte überträgt, sind es ihr Verhalten und ihre Handlungen, die bestimmt werden müssen. Eine nützliche und kostenminimale Aufstellung, um dies zu bewerkstelligen, ist die Benutzung eines Master-Slave-Kamerasystems, bestehend aus einer Übersichtskamera mit Weitwinkeloptik und mindestens einer aktiven Kamera. Die Absicht eine Szene mit möglichst wenig Kameras zu observieren bedingt eine intelligente Aufmerksamkeitssteuerung, die in der Lage sein muss, möglichst alle interessanten Situationen in der Szene nach bestimmten Kriterien zu beobachten. Am Fraunhofer IOSB Ettlingen sind mehrere Master-Slave-Kamerasysteme im Einsatz, siehe z.B. [1]. In dieser Studienarbeit soll darauf aufbauend eine intelligente Aufmerksamkeitssteuerung für aktive Kameras entwickelt werden, um die Menge der aus der Szene gewonnenen Information zu maximieren, z.B. durch Nahaufnahmen von Personen und Objekten, um diese zu identifizieren oder um detaillierte Ansichten mehrerer miteinander interagierender Personen zu deuten. Ein essenzieller Teil der Arbeit ist die Erfassung spezifischer Informationen der Kameras, um Aussagen über die zeitlichen Verzögerungen der Kamerabewegungen treffen zu können, um die Kosten abschätzen zu können, die bei Bewegung der Kamera an eine gewünschte Position anfallen. Wir erwarten durch Einbindung dieser Informationen in den Optimierungsprozess bei der Steuerung der Kameras verlässliche und robuste Aufnahmen interessanter Situationen in der Szene. Mehrere bereits vorhandene Arbeiten zur Steuerung aktiver Kameras sollen miteinander verglichen werden, wovon die passendste nachimplementiert und erweitert werden soll. Verwandte Arbeiten von Interesse sind unter anderem “Scheduling an active camera to observe people” [2] und “Systems and algorithms for autonomous and scalable crowd surveillance using robotic PTZ cameras assisted by a wide-angle camera” [3].

Inhaltsverzeichnis

1	Einleitung	1
1.1	Motivation	1
1.2	Problemerkfassung	2
1.3	Gliederung der Arbeit	2
2	Verwandte Arbeiten	3
3	Grundlagen	6
3.1	Mehrkamerasystem	6
3.2	Scheduling	8
3.2.1	Detektion, Markierung, Verfolgung, Prediktion	8
3.2.2	Generierung und Priorisierung von Observierungsanfragen	10
3.2.3	Auswahl der Kameraparameter (p-frame selection)	11
4	Entwurf	16
4.1	Temporale Distanz	16
4.2	Aufmerksamkeitskontrolle	19
4.2.1	Verkürzung der Abfangzeit	19
4.2.2	Auswahl der Kamerabewegungen (p-move selection)	20
4.2.3	Mikroprädiktion und Feinverstellung	21
4.2.4	Einfluss des Zooms auf die Bewertung von Kamerabewegungen	21
5	Auswertung	24
5.1	Simulation	24
5.2	Statistiken	24
5.3	Beispielhafter Auszug aus der Simulation	31
6	Zusammenfassung	34
6.1	Ergebnis	34
6.2	Ausblick	34
A	Abbildungsverzeichnis	35
B	Literaturverzeichnis	36

1 Einleitung

1.1 Motivation

In einer zunehmend dichter vernetzten Welt, in der Sensoren riesige Mengen an Daten durch die automatisierte Erfassung ihrer Umwelt generieren, wird es immer wichtiger für die nachgeschalteten informationsverarbeitenden Systeme, den Fokus der Sensorik auf die - je nach Anwendungsfall - für sie wesentlichen Bereiche zu richten.

Im Speziellen soll es in dieser Studienarbeit darum gehen, auf sinnvolle Art und Weise Bildmaterial hoher Auflösung von Personen und Objekten zu generieren, die in zeitlich unbestimmten Abständen an den Rändern des Bildbereiches einer Videokamera mit starrem und weitem Blickfeld erscheinen und diesen in beliebiger Form und Geschwindigkeit durchqueren.

Da eine Videokamera mit Weitwinkeloptik starrem Blickfeld für gewöhnlich nur Bildmaterial mit relativ geringer Auflösung generieren kann, wird sie lediglich dazu verwendet, einzelne Personen zu detektieren und zu verfolgen. Sie soll Daten erfassen, die nun der Koordination mehrerer aktiver Videokameras dienen, die in der Lage sind, durch senkrechtes und waagerechtes Schwenken, ihren Bildbereich zu verstellen und durch das Einstellen unterschiedlicher Zoomstufen Bildmaterial höherer Auflösung zu generieren. Wir nennen diese Art von Kameras im Weiteren PTZ-Kameras.

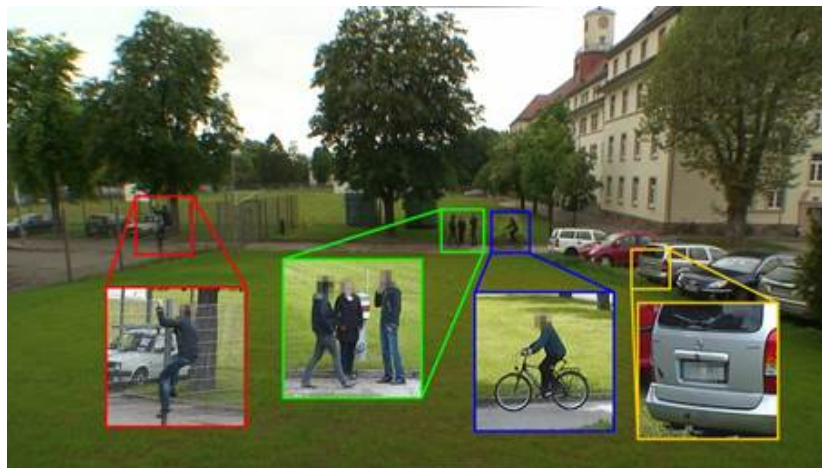


Abbildung 1: Übersichtsbild mit Nahaufnahme von Personen und Objekten.

1.2 Problemerkfassung

Da oft die Anzahl zu beobachtender Personen die der zur Verfügung stehenden PTZ-Videokameras übersteigt, ist es nötig, die Ressourcen der einzelnen Videokameras sinnvoll einzuteilen. Bei der Nutzung verschiedener Vergrößerungsstufen gilt es, einen Kompromiss zwischen dem sichtbaren Bildbereich und der Auflösung des Bildmaterials zu finden. Außerdem gilt es, die Zeit zu berücksichtigen, die dazu nötig ist, die momentane Konfiguration einer Videokamera in eine andere zu überführen.

Bedingt durch die Dynamik in der Szene und die zur Zuteilung der Personen an die PTZ-Videokameras benötigten Zeit sind außerdem eine Positionsabschätzung und Mikroprädiktionen nötig.

Beim Anfahren der einzelnen Personen von Interesse gibt es je nach Anwendungsfall verschiedene Kriterien, die berücksichtigt werden müssen. Erstens müssen die verschiedenen Personen priorisiert werden. Das kann auf verschiedene Art und Weise erfolgen, z.B. durch ihre aktuelle Größe (ist die Person groß genug, um identifiziert werden zu können) oder ihre Position und Richtung (wird die Person den Sichtbereich in unmittelbarer Zeit verlassen) oder ob bereits genügend hochaufgelöste Bilder zu dieser Person vorliegen. Zweitens muss darüber entschieden werden, wie der Nutzen durch das Anfahren einzelner Personen, gemessen werden kann. Und drittens muss dieser Nutzen maximiert werden.

1.3 Gliederung der Arbeit

Diese Studienarbeit stellt im folgenden Kapitel verwandte Arbeiten auf dem Gebiet der Aufmerksamkeitssteuerung von Videonetzwerken vor. Die theoretischen Grundlagen, die für das Betreiben eines Master-Slave Mehrkameranensystems nötig sind, werden in Kapitel 3.1 vorgestellt. Anschließend wird in Kapitel 3.2 der dieser Arbeit zugrundeliegende Ansatz [3] von Xu et al. vorgestellt, um diesen schließlich in Kapitel 4 um eine zeitliche Komponente zu erweitern, die es nun zusätzlich ermöglicht, die Zeit der Kamera sich von einem Punkt zu einem anderen zu bewegen, zu berücksichtigen. In Kapitel 5 werden das ursprüngliche Verfahren [3] und die hier erarbeiteten Anpassungen und Erweiterungen auf künstlich erzeugten Daten quantitativ ausgewertet, gefolgt von einer Zusammenfassung und dem Ausblick.

2 Verwandte Arbeiten

Es existieren bereits eine Vielzahl von Arbeiten zur Thematik der Aufmerksamkeitssteuerung. Einige davon sollen in diesem Kapitel erwähnt werden.

„Towards on-line saccade planning for high-resolution image sensing“ stellt eine Methode vor, die versucht vorausschauend eine optimale Reihenfolge zu bestimmen, in der eine PTZ-Kamera mehrere bewegliche Ziele anfahren sollte, um die Leerlaufzeiten beim Schwenken von einem Ziel zum anderen, zu minimieren. Es werden eine Master-Kamera für einen Überblick über die Szene und eine aktive PTZ-Kamera verwendet. Der Ansatz lässt sich jedoch nur schwer auf mehrere PTZ-Kameras anwenden.

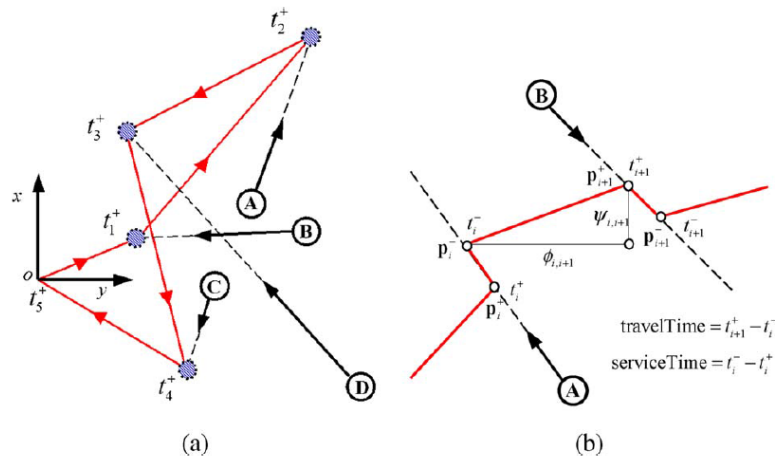


Abbildung 2: Veranschaulichung, wie eine PTZ-Kamera mehrere bewegte Ziele anfahren könnte, um Leerlaufzeiten zu minimieren.

„Planning Ahead for PTZ Camera Assignment and Handoff“ stellt ein Verfahren vor, das mehrere Weitwinkeloptik-Kameras und mehrere PTZ-Kameras kollaborieren lässt. Die Übersichtskameras markieren einzelne Personen von Interesse, die dann von jeweils einer aktiven PTZ-Kamera verfolgt werden. Die Bildbereiche der PTZ-Kameras können sich zwar überschneiden, sind im allgemeinen aber unterschiedlich, so dass das hier beschriebene Verfahren die Zuweisung von Personen an PTZ-Kameras dahingehend optimiert, dass es vorausschauend plant, welche Bereiche eine Person voraussichtlich durchschreiten wird und welche PTZ-Kameras in der Lage sein werden, sie zu observieren, sofern sie nicht bereits mit der Observierung einer anderen Person beauftragt sind. Bei Übergang einer Person vom Bildbereich einer PTZ-Kamera in den einer anderen, findet eine Weitergabe des Observierungsauftrags statt.

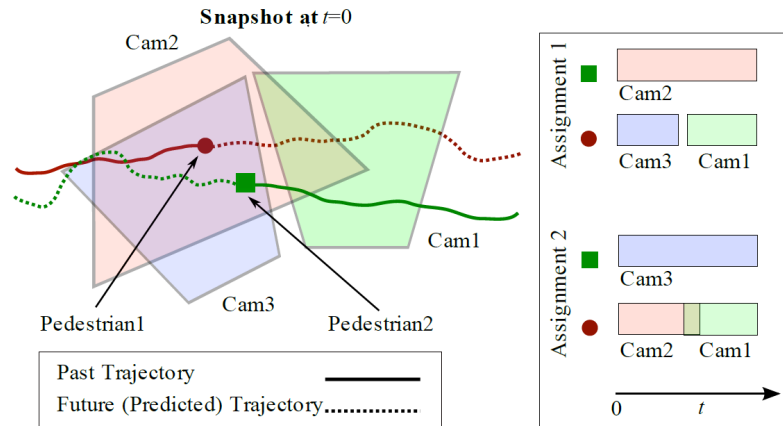


Abbildung 3: Veranschaulichung, wie Personen Zuständigkeitsbereiche einzelner PTZ-Kameras durchqueren und eine optimale vorausschauende Zuweisung an einzelne PTZ-Kameras für eine möglichst lückenfreie Observierung stattfindet.

„Surveillance Camera Coordination Through Distributed Scheduling“ bearbeitet eine ähnliche Problemstellung wie [5]. Auch hier durchqueren Personen Bereiche, die nur von einzelnen PTZ-Kameras beobachtet werden können. Im besonderen wird in dieser Arbeit versucht, die Komplexität der Vorausplanung, welche Kamera welche Person observieren sollte, dadurch zu vermindern, dass eine Entscheidungsfindung auf möglichst lokaler Ebene zwischen benachbarten PTZ-Kameras stattfindet.

„Systems and Algorithms for Autonomous and Scalable Crowd Surveillance Using Robotic PTZ Cameras Assisted by a Wide-Angle Camera“ verfolgt den Ansatz mit Hilfe einer fixen Kamera mit Weitwinkeloptik eine Szene grob zu erfassen, um aus dieser so gewonnenen Übersicht über die Szene global eine optimale Zuweisung von Personen an mehrere aktive PTZ-Kameras zu bewerkstelligen. Für die Bestimmung dieser globalen Lösung werden verschiedene Attribute der Personen erfasst, wie zum Beispiel die Zeit, die noch verbleibt, bis eine Person die Szene verlässt. Im Unterschied zu den bereits erwähnten Verfahren, wird hier auch ermöglicht, dass eine PTZ-Kamera mehrere Personen auf einmal observieren kann, sofern diese nahe genug beieinander stehen. Da dieser Ansatz von Xu et. al. am vielversprechendsten erscheint, wird er im Rahmen dieser Studienarbeit als Grundlage für die Entwicklung weiterer Verfahren dienen.

„Exact Algorithms for Non-Overlapping 2-Frame Problem with Non-Partial Coverage for Networked Robotic Cameras“ ist der Versuch von Xu et al. einen

großen Nachteil des in [3] beschriebenen Verfahrens zu umgehen. Da das Optimierungsproblem zur Findung einer bestmöglichen Zuweisung von mehreren Personen an mehrere PTZ-Kameras zu komplex und rechenaufwendig ist, begrenzt man sich in dieser Arbeit auf nur noch zwei PTZ-Kameras und ist dadurch in der Lage effektivere Algorithmen zu formulieren, die einen Aufwand von $O(n^2)$ bis $O(n^3)$ erfordern.

3 Grundlagen

Die in dieser Studienarbeit entwickelten Verfahren setzen eine Kameraarchitektur, welche das Zusammenwirken und die Koordination mehrerer PTZ-Kameras ermöglicht, und teilweise bereits bestehende Arbeiten aus dem Gebiet der automatisierten „crowd surveillance“ voraus.

Deshalb werden in diesem Kapitel die Arbeiten „Automatische Konfiguration eines Mehrkameranystems zur Verfolgung von Objekten in unstrukturierten Umgebungen“ und „Systems and Algorithms for Autonomous and Scalable Crowd Surveillance Using Robotic PTZ Cameras Assisted by a Wide-angle Camera“ möglichst in nur notwendiger Detailtiefe erläutert und anschließend in Kapitel 4 wieder aufgegriffen und erweitert.

3.1 Mehrkameranystem

Um die Verwirklichung einer sinnvollen Aufmerksamkeitssteuerung gewährleisten zu können, ist es nötig, Bildpunkte aus den 2D-Koordinatensystemen der Bildbereiche der einzelnen Videokameras zueinander zuzuordnen zu können.

Eine Möglichkeit, dies zu bewerkstelligen, wird in [1] wie folgt beschrieben:

Das Gesamtsystem besteht aus $(p + 1)$ Videokameras. Hiervon fungiert eine als Master-Kamera, deren statisches Blickfeld als Referenz-Bild für die restlichen (Slave-)Videokameras dient. Die restlichen p Videokameras haben die zusätzliche Funktionalität, ihr Blickfeld durch Verstellen ihrer „pan“, „tilt“ und „zoom“-Parameter zu verändern. Diese Art von Kameras werden im weiteren Verlauf der Studienarbeit als PTZ-Kameras bezeichnet.

Durch das Wiederfinden von charakteristischen Merkmalpunkten aus dem Bildbereich der Master-Kamera in den Bildbereichen der einzelnen PTZ-Kameras wird stichprobenartig eine Zuordnungsvorschrift von Master-Koordinaten zu PTZ-Kamera-Koordinaten geschaffen. Dies geschieht für jede einzelne PTZ-Kamera separat. Anschließend wird mit Hilfe von linearer Regression eine flächendeckende Zuordnungsfunktion angenähert, welche nun zu beliebigen Master-Kamera-Koordinaten die entsprechenden PTZ-Kamera-Koordinaten liefert.

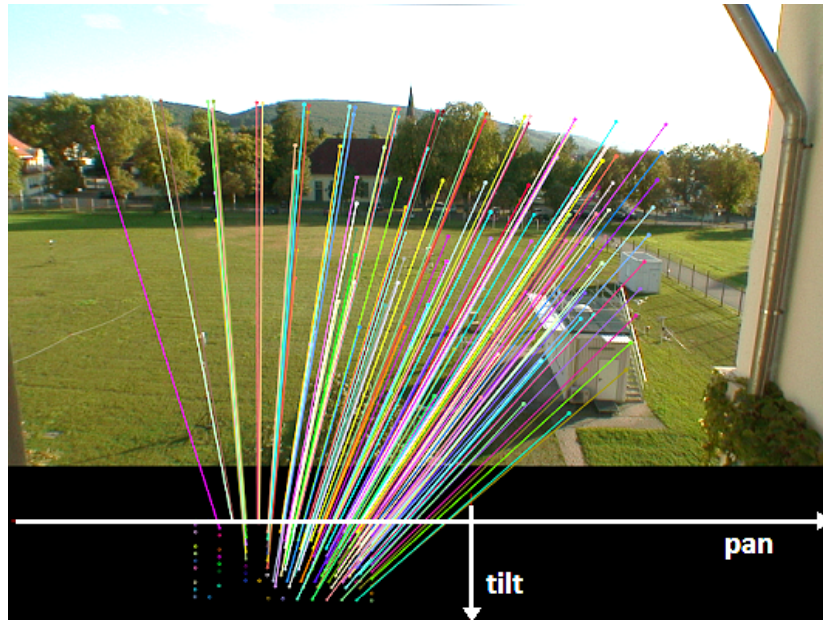


Abbildung 4: Stützpunkte der Abbildung zwischen Master-Koordinaten und PTZ-Koordinaten.

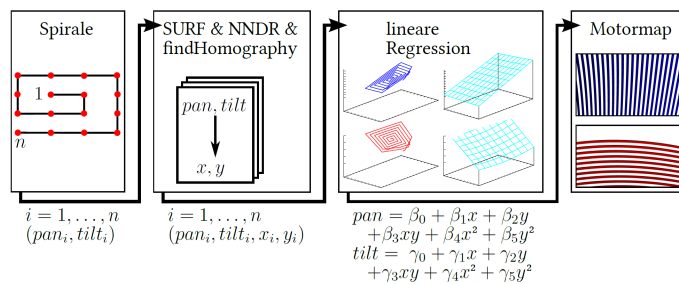


Abbildung 5: Zuordnungsprozess von Kamerakoordinaten zueinander. a) Übereinstimmungen finden b) einander zuordnen c) lineare Regression d) fertige Motormap.

Im Weiteren wird, falls nicht anders gekennzeichnet, nur noch von Master-Kamera-Koordinaten gesprochen. Zu beachten ist, dass aufgrund der unterschiedlichen Positionierung der PTZ-Kameras in der realen Welt, unterschiedliche Blickwinkel auf die im Master-Bild erkannten Merkmalpunkte vorliegen. Daraus folgt, dass die von einer PTZ-Kamera zurückzulegende temporale Distanz, um von der aktuellen Position (u, v) auf eine andere Position (u', v') umzuschwenken, im Allgemeinen nicht bei allen PTZ-Kameras gleich groß sein

muss und gerade bei den hier verwendeten einfachen Standardkameras besonders zu betrachten ist, weil diese sich langsamer und ungenauer bewegen als hochtechnologische Geräte. Diese Eigenschaft wird später bei der Berechnung optimaler Strategien der Blick-Kontrolle in Kapitel 4 berücksichtigt werden.

3.2 Scheduling

Die von „Xu et al.“ [3] vorgestellte Arbeit liefert ein Verfahren, das in der Lage ist, auch bei sehr vielen Personen in der zu beobachtenden Szene, gute Ergebnisse zu liefern. Das Verfahren nutzt eine Videokamera mit weitem Blickfeld, um eine Übersicht über die Szene zu erhalten. In der von der Übersichtskamera erzeugten Bildfolge werden nun mit Hilfe eines geeigneten Detektors (z.B. Blob-Detektor) sich bewegende Objekte detektiert, mit einer eindeutigen Bezeichnung (ID) markiert und mit Hilfe eines Kalman-basierten Trackers durch den Bildstrom hindurch verfolgt. Aus den nun bekannten Positionen der einzelnen Fußgänger werden Observierungsanfragen generiert und an die Software-Komponente zur Aufmerksamkeitssteuerung übergeben. Diese untersucht anhand der Observierungsanfragen alle möglichen Kamerakonfigurationen und die damit verbundenen Bildbereiche für die PTZ-Kameras auf ihre Güte hin, hinsichtlich der Frage, wie viele Observierungsanfragen gänzlich im Bildbereich der PTZ-Kamera eingeschlossen sind, und wie hoch die Auflösung dieses Bildes ist. Anschließend wird ein Optimierungsproblem definiert und die Sättigung, bei der Auswahl von p Konfigurationen, maximiert. Nachdem eine PTZ-Kamera die angeforderte Konfiguration erreicht und somit die Observierung eingeleitet hat, verfolgt sie ihr Ziel selbstständig so lange durch sogenannte Mikroprädiktionen und Feinverstellung, bis neue Anweisungen vorliegen.

Dieses Verfahren schneidet im Vergleich zu anderen einfachen Scheduling-Verfahren, wie z.B. „earliest deadline first“ (EDF) und „first come first serve“ (FCFS) besser ab, berücksichtigt jedoch die zeitliche Distanz nicht, die von den PTZ-Kameras zurückgelegt werden muss, um die geforderte Konfiguration einzustellen. Während dieser Zeitabschnitte liefern die Kameras kein verwertbares Bildmaterial, was zu der Annahme führt, dass an dieser Stelle weitere Optimierungen vorgenommen werden könnten. Hierzu mehr in Kapitel 4. Weitere Details und Formalisierungen werden in den folgenden Unterkapiteln vorgestellt.

3.2.1 Detektion, Markierung, Verfolgung, Prediktion

Wie bereits erwähnt, stützt sich das in [3] vorgestellte Verfahren auf zuverlässige Ergebnisse von Detektoren, Markern und Trackern. Sie dienen dazu, relevante Daten zu Personen möglichst akkurat zu bestimmen. Hierbei geht es hauptsächlich darum eine Person in Bildfolgen wiederzuerkennen, und ihre Position, Ge-

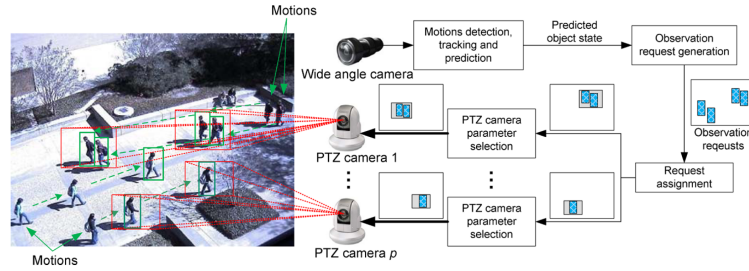


Abbildung 6: Die einzelnen Schritte im Verfahren [3].

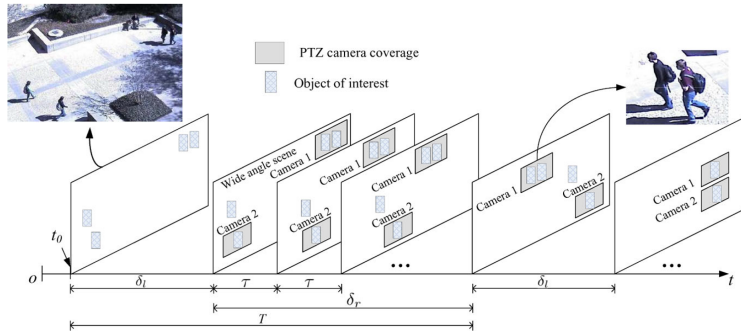


Abbildung 7: Übersicht zum zeitlichen Ablauf des Scheduling bei Xu et al. δ_l bezeichnet die Abfangzeit zur Positionierung der Kameras, δ_r ist die Aufnahmezeit.

schwindigkeit und Ausmaße zu erfassen.

Die Größe ergibt sich aus der Breite b und der Höhe h des Rechtecks, welches, mit Kanten orthogonal zu den Koordinaten-Achsen, die Person minimal und vollständig beinhaltet. Die Position der Personen ist gleichzusetzen dem Zentrum dieses Rechtecks mit den Koordinaten (x, y) . Die senkrechten und waagerechten Geschwindigkeiten \dot{x}, \dot{y} lassen sich aus der zurückgelegten Strecke der Position der Person in einem bestimmten Zeitintervall mit Länge t_{int} bestimmen. Das Ausmaß der Personen wird definiert über die Weite w und die Höhe h des Rechtecks.

Der Zustand einer Person p_i kann zum Zeitpunkt t somit beschrieben werden durch:

$$p_i(t) = [x(t), y(t), h(t), b(t), \dot{x}(t), \dot{y}(t)]^T. \quad (1)$$

Würde man nun mit Hilfe dieser Informationen versuchen eine optimale Menge von Kamerakonfigurationen zu bestimmen, welche möglichst zufriedenstellend die Observierung der Personen bewerkstelligen sollen, so wären die Personen während der Zeit, die zur Berechnung eben jener Konfigurationen und der Zeit, die die PTZ-Kameras zum Umkonfigurieren benötigen, längst nicht mehr an den der Berechnung zugrundeliegenden Positionen. Daher ist es sinnvoll, durch die Verwendung von z.B. Kalman-Filtern den Zustand der Personen in naher Zukunft abzuschätzen. Diese Zeitspanne, die für Rechenzeit und die Umkonfiguration der Kameras reserviert ist, nennen wir Abfangzeit. Der Zustand einer Person zu einer zukünftigen Zeit lässt sich beschreiben durch:

$$\hat{p}_i(t) = [\hat{x}_i(t), \hat{y}_i(t), \hat{h}_i(t), \hat{b}_i(t), \hat{\dot{x}}_i(t), \hat{\dot{y}}_i(t)]^T. \quad (2)$$

Diese Informationen werden an die nächste Software-Komponente zur Generierung und Priorisierung von Observierungsanfragen weitergereicht.

3.2.2 Generierung und Priorisierung von Observierungsanfragen

Für die weitere Verwertung der Personen-Zustände, werden diese um Informationen ergänzt, während Attribute, die für folgende Komponenten keine Rolle mehr spielen weggelassen werden:

$$r_i(t) = [\hat{x}_i(t), \hat{y}_i(t), \hat{h}_i(t), \hat{b}_i(t), z_i(t), \omega_i(t)]^T. \quad (3)$$

$$R^n(t) = \begin{bmatrix} r_1(t) \\ r_2(t) \\ \dots \\ r_n(t) \end{bmatrix} \quad (4)$$

Hierbei sind z_i und ω_i Hilfsvariablen, die der späteren Findung von guten PTZ-Kamera Konfigurationen dienen. z_i gibt an, wie der Zoom einer PTZ-Kamera eingestellt sein müsste, um die Person bei höchstmöglicher Auflösung noch vollständig erfassen zu können. ω_i ist ein Wert zwischen 0 und 1, der angibt, wie dringlich es ist, diese Personen zu filmen. Er ergibt sich daraus, wieviel Bildmaterial zu dieser Person bereits generiert wurde, und wieviel Zeit voraussichtlich noch verbleibt, bis sie den sichtbaren Gesamtbildbereich verlässt.

Es gilt:

$z_i \in [0, 1]$, wobei 0 der minimalen und 1 der maximalen Zoomstufe, die den gesamten Bildbereich einschließt, entsprechen.

$\omega_i \in [0, 1]$, wobei

$$\omega_i(t) = \mu_i(t) \cdot \nu_i(t). \quad (5)$$

$\mu_i(t)$ stellt die Dringlichkeit im Bezug auf die Zeit dar, die noch bleibt, um den Passanten zu observieren.

$$\mu_i(t) = \rho^{\hat{d}_i - t}. \quad (6)$$

\hat{d}_i ist der Zeitpunkt, an dem der Passant, bei gleichbleibender momentaner Geschwindigkeit, die Szene verlassen würde. $\rho \in [0, 1]$ ist ein Parameter zur Gewichtung dieser Eigenschaft.

$\nu_i(t)$ zeigt die Dringlichkeit eine Person zu observieren an, hinsichtlich des bereits in der Vergangenheit von ihm erstellten Bildmaterials $\zeta_i(t)$ (verlangter Zoom dividiert durch den von der die Person beobachtenden PTZ-Kamera eingestellten Zoom zum Zeitpunkt t) und des verlangten Bildmaterials n_e .

$$\nu_i(t) = \max \left(1 - \frac{\eta_i(t)}{n_e} \right). \quad (7)$$

$$\eta_i(t) = \sum_{t_k < t} \zeta_i(t_k). \quad (8)$$

Weitere und andere Bewertungsfunktionen können je nach Anwendungsfall sinnvoll sein und werden in Kapitel 4 vorgestellt.

3.2.3 Auswahl der Kameraparameter (p-frame selection)

Aufgabe der Komponente zur Aufmerksamkeitssteuerung ist es nun aus der Menge der priorisierten Observierungs-Anfragen eine möglichst optimale Konfiguration für die PTZ-Kamera-Parameter zu bestimmen.

Eine PTZ-Kamera-Konfiguration (p, t, z) besteht aus pan p und tilt t Koordinaten zur Ausrichtung an den Koordinatenachsen, und einem Zoomfaktor z , der festlegt, wie groß der zu beobachtende Bildbereich und die Auflösung dessen sind. Da in Kapitel 3.1 eine Transformationsvorschrift zwischen pan- und tilt-Werten der einzelnen PTZ-Kameras und den x - und y -Koordinaten im Bildbereich der Übersichtskamera erarbeitet wurde, verwenden wir auch für die PTZ-Konfigurationen einheitlich Koordinatenangaben der Übersichtskamera.

$$c = [x, y, z]^T. \quad (9)$$

C^m bezeichnet die Menge aller möglichen Kameraparameter-Belegungen.

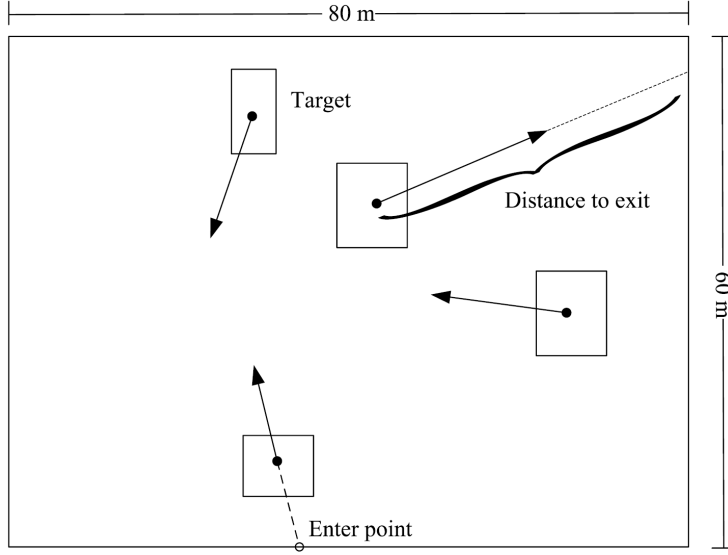


Abbildung 8: Personen und ihre Bewegungsvektoren. Markierung der Zeit, bis eine Person den Bildbereich verlässt.

$$C^m = \{c_i | c_i = [x_i, y_i, z_i]\} \quad (10)$$

Die Anzahl möglicher Kamerakonfigurationen soll begrenzt sein. ($|C^m| = m$) Dies erleichtert das Entwerfen von Algorithmen und senkt die zur Bestimmung optimaler Teilmengen benötigte Rechenzeit enorm. Deshalb wird in [3] der Raum aller möglichen Kamerakonfigurationen durch eine Diskretisierung der pan-, tilt- und Zoomwerte vereinfacht.

Als nächstes müssen aus C^m entsprechend der Anzahl an PTZ-Kameras p Konfigurationen ausgewählt werden. Da im Normalfall die Anzahl der Observierungsanfragen die der zur Verfügung stehenden PTZ-Kameras übersteigt, können nicht immer alle Anfragen erfüllt werden. Auch stehen meist mehrere Kamerakonfigurationen zur Auswahl, die dieselbe Anfrage abdecken können. Hier wird auf die zuvor vollzogene Priorisierung der einzelnen Observierungsanfragen zurückgegriffen.

Die Güte einer Kamerakonfiguration wird definiert über eine Sättigungsfunktion:

$$s(c, r_i(t)) = \omega_i(t) \cdot I(c, r_i(t)) \cdot \min\left(\frac{z_i}{z}, 1\right). \quad (11)$$

Zur Erinnerung: z_i ist der kleinstmögliche Zoom, den eine PTZ-Kamera einstellen kann, um die Observierungsanfrage r_i noch vollständig im Bildbereich zu umschließen. z wiederum ist der tatsächliche Zoom, der Kamerakonfiguration c .

$$I(c, r_i(t)) = \begin{cases} 1, & \text{falls } r_i(t) \text{ vollständig in } c \text{ liegt} \\ 0, & \text{sonst.} \end{cases} \quad (12)$$

Die Sättigung einer Kamerakonfiguration im Bezug auf alle vorliegenden Observierungsanfragen ist gegeben durch:

$$s(c, R^n(t)) = \sum_{i=1}^n s(c, r_i(t)). \quad (13)$$

Die Sättigung einer Menge von Kamerakonfigurationen hinsichtlich aller vorliegenden Observierungsanfragen ist:

$$s(C^p, R^n(t)) = \sum_{i=1}^p \sum_{j=1}^n s(c_i, r_j(t)). \quad (14)$$

Mit Hilfe der letzten Funktion wird ein Maximierungsproblem definiert, dessen Lösung eine Auswahl von p zum Zeitpunkt t optimalen Kamerakonfigurationen liefert:

$$C^{p*}(t) = \arg \max_{C^p} s(C^p, R^n(t)) \quad (15)$$

$$C^{p*}(t) = \begin{bmatrix} c_1^* \\ c_2^* \\ \dots \\ c_p^* \end{bmatrix} \quad (16)$$

An die möglichen Vektoren C^{p*} wird die einschränkende Bedingung gestellt, dass je zwei Konfigurationen nicht dazu führen dürfen, dass eine Observierungsanfrage durch beide Konfigurationen gleichzeitig erfüllt wird. Der Informationsgewinn wäre dadurch nicht viel höher, da beide generierten Bilder etwa dasselbe anzeigen würden. Diese Bedingung wird non-overlapping-condition (NOC) genannt.

Xu et al. stellen in [3] einen Algorithmus vor, dessen Laufzeit in $O(n/\epsilon^3 + p^2/\epsilon^6)$ liegt, um dieses Maximierungsproblem bei Einhaltung der NOC zu lösen.

Die Aufmerksamkeitssteuerung ist somit eine Funktion, die zu einem Vektor aus Observierungsanfragen mit Länge n einen Vektor aus PTZ-Kamera-Konfigurationen mit Länge p aus der Menge aller möglichen Konfigurationen generiert, wobei n die Anzahl aller Observierungs-Anfragen und p die Anzahl der vorhandenen

PTZ-Kameras angibt:

$$f_{gc}(R^n(t)) = C^{p^*}(t). \quad (17)$$

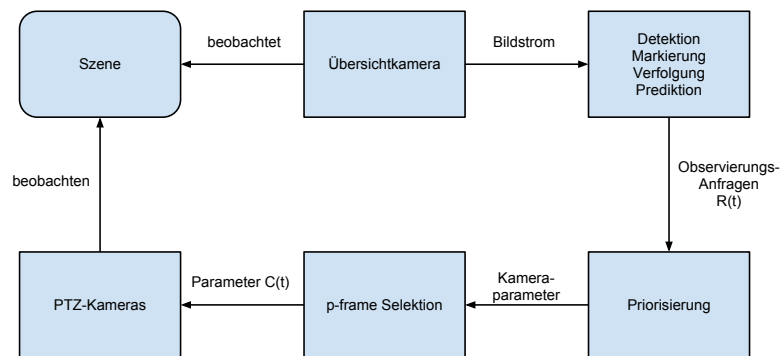


Abbildung 9: Abstrakte Übersicht zum Ablauf des Scheduling nach Xu.

Algorithm 1 Lattice-based Algorithm

```

1 begin
2   for  $j \leftarrow 1$  to  $|L|$  do  $O(1/\epsilon^3)$ 
3      $l[j] = s(\tilde{c}_j)$   $O(n)$ 
4      $\tilde{Q}_0(\tilde{c}_j) = \emptyset;$   $O(1)$ 
5      $s(\tilde{Q}_0(\tilde{c}_j)) = 0;$   $O(1)$ 
6   end
7   for  $k \leftarrow 1$  to  $p$  do  $O(p)$ 
8      $\tilde{C}^{k*} = \emptyset;$   $O(1)$ 
9      $s(\tilde{C}^{k*}) = 0;$   $O(1)$ 
10    for  $u \leftarrow 1$  to  $|L|$  do update  $\tilde{C}^{k*}, O(1/\epsilon^3)$ 
11      if  $s(\tilde{C}^{k*}) < s(\tilde{Q}_{k-1}(\tilde{c}_u)) + l[u]$ 
12        then
13           $\tilde{C}^{k*} = \tilde{Q}_{k-1}(\tilde{c}_u) \cup \{\tilde{c}_u\};$   $O(1)$ 
14           $s(\tilde{C}^{k*}) = s(\tilde{Q}_{k-1}(\tilde{c}_u)) + l[u];$   $O(1)$ 
15        end
16      end
17    for  $u \leftarrow 1$  to  $|L|$  do update  $\tilde{Q}_k(\tilde{c}_u), O(1/\epsilon^3)$ 
18       $\tilde{Q}_k(\tilde{c}_u) = \tilde{Q}_{k-1}(\tilde{c}_u) \cup \emptyset;$   $O(1)$ 
19       $s(\tilde{Q}_k(\tilde{c}_u)) = s(\tilde{Q}_{k-1}(\tilde{c}_u));$   $O(1)$ 
20      for  $v \leftarrow 1$  to  $|L|$  do  $O(1/\epsilon^3)$ 
21        if  $s(\tilde{Q}_k(\tilde{c}_u)) < s(\tilde{Q}_{k-1}(\tilde{c}_v)) + l[v]$  AND  $\{\tilde{c}_u, \tilde{c}_v\} \cup \tilde{Q}_{k-1}(\tilde{c}_v)$ 
22          satisfies the VNOC  $O(p)$ 
23        then
24           $\tilde{Q}_k(\tilde{c}_u) = \tilde{Q}_{k-1}(\tilde{c}_v) \cup \{\tilde{c}_v\};$   $O(1)$ 
25           $s(\tilde{Q}_k(\tilde{c}_u)) = s(\tilde{Q}_{k-1}(\tilde{c}_v)) + l[v];$   $O(1)$ 
26        end
27      end
28    end
29  return  $\tilde{C}^{p*};$ 
30 end

```

Abbildung 10: Algorithmus von Xu et al. zur Bestimmung einer Menge optimaler Kamerakonfigurationen.

4 Entwurf

Das in Kapitel 3.2 vorgestellte Verfahren [3] weist den gegebenen Observierungsanfragen quantitative und dynamische Prioritäten $\omega_i \in [0, 1]$ zu. Berücksichtigt wird hierdurch jedoch nur, wie oft eine Person bei welcher PTZ Kameraauflösung bereits beobachtet wurde, und wieviel Zeit noch verbleibt, bis der Passant den beobachtbaren Bildbereich der Übersichts-Kamera verlässt.

Diese Priorisierungsfunktionen sind sinnvoll und sollen daher beibehalten werden. Es scheint allerdings möglich, dieses bestehende Verfahren noch weiter zu verbessern, indem man berücksichtigt, wie viel Zeit die PTZ-Kameras benötigen, um ihre aktuellen Parameter-Einstellungen in die von der Aufmerksamkeitssteuerung vorgegebenen zu überführen. Durch verkürzte Schwenk- und Zoomzeiten würden die PTZ-Kameras mehr Bildmaterial generieren.

Zunächst sollen daher die Überführungsgeschwindigkeiten der Parameter erfasst werden. Anschließend sollen neben den Observierungsanforderungen $R(t)$ auch noch die Kamerazustände $A(t)$ erfasst werden. Dadurch kann eine (temporale) Distanzmetrik definiert werden, die in die Beurteilung der Kameraparameter auf ihre Güte hin, benutzt werden kann. Anstelle der p optimalen Kamerakonfigurationen $C(t)$ werden p optimale Kamerabewegungen ausgewählt. Die folgenden Abschnitte beschreiben das Vorgehen.

4.1 Temporale Distanz

Eine PTZ-Kamera benötigt eine gewisse Zeit, um ihre momentane Parameter-Konfiguration für pan-, tilt- und zoom-Wert $c = [p, t, z]^T$ in eine andere $c' = [p', t', z']^T$ zu überführen (Angaben in Koordinaten der Übersichtskamera). Diese hängt von den hardwarebedingten Schwenk- und Zoomgeschwindigkeiten ab und sollte bei Kameras gleicher Bauart gleich sein, ist es in der Praxis aber leider nicht. Durch die in Kapitel 3.1 vorgegebene Vereinheitlichung aller Koordinatenangaben wird die Parameterbelegung der einzelnen PTZ-Kamera Konfigurationen in Bildkoordinaten (u, v) des Master-Koordinatensystems angegeben.

$$c = [x, y, z]^T \quad (18)$$

Durch die unterschiedliche Aufstellung und Ausrichtung im Raum der einzelnen PTZ-Kameras folgt, dass die zum Umkonfigurieren benötigte Zeit τ für die PTZ-Kameras i, j im Allgemeinen unterschiedlich sind.

$$i \neq j \Rightarrow \tau_i(c, c') \neq \tau_j(c, c') \quad (19)$$

Es empfiehlt sich, die Geschwindigkeiten experimentell zu bestimmen. Die PTZ-Kameras, die für diese Studienarbeit verwendet wurden, zeigten - trotz gleicher Bauart - leicht unterschiedliche Schwenk- und Zoomgeschwindigkeiten, die sich

recht gut linear annähern lassen. Siehe Abbildung 9 für die Schwenkgeschwindigkeit und Abbildung 10 die Zoomgeschwindigkeit für eine der in dieser Arbeit verwendeten Kameras. Es stellte sich heraus, dass hauptsächlich das Verstellen der Zoomstufen für Leerlaufzeiten, in denen die Kameras kein Bildmaterial liefern, verantwortlich waren.

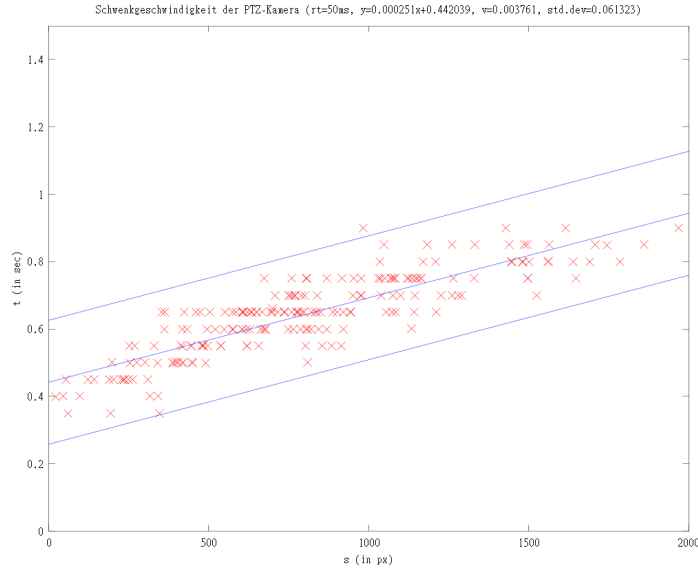


Abbildung 11: Schwenkgeschwindigkeiten der PTZ-Kamera. Auf der x -Achse die zurückgelegte Strecke in px, auf der y -Achse die benötigte Zeit in sek.

Die experimentell bestimmten Werte mit einbezogen, ergibt sich der momentane Zustand einer PTZ-Kamera a_i als:

$$a_i(t) = [x_i(t), y_i(t), z_i(t), v_i^{pan,tilt}, v_i^{zoom}]^T \quad (20)$$

x_i , y_i und z_i geben die momentane Ausrichtung einer PTZ-Kamera an den Koordinatenachsen und dessen Zoom an. $v_i^{pan,tilt}$ gibt an, wieviel Strecke im Bild bzgl. der Koordinaten (u, v) die PTZ-Kamera pro Sekunde zurücklegen kann, während v_i^{zoom} angibt, wie stark der Zoom $z_i \in [0, 1]$ pro Sekunde verstellt werden kann.

Die temporale Distanz zwischen einer real vorliegenden PTZ-Kamera Konfiguration $a_j(t)$ und einer Wunschkonfiguration $c_i = [p, t, z]^T$ sei definiert als:

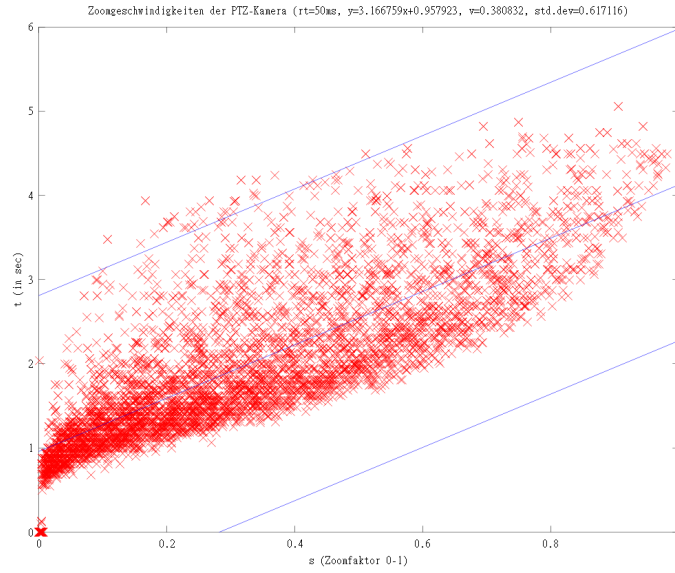


Abbildung 12: Auf der x -Achse die Verstellung des Zooms zw. 0 und 1, auf der y -Achse die benötigte Zeit in sek.

$$\tau(c_i, a_j(t)) = \max \left(\frac{\sqrt{(x_i - x_j(t))^2 + (y_i - y_j(t))^2}}{v_j^{pan,tilt}}, \frac{|z_i - z_j(t)|}{v_j^{zoom}} \right) \quad (21)$$

4.2 Aufmerksamkeitskontrolle

Neben den Observationsanfragen liegen der Aufmerksamkeitssteuerung auch die aktuellen Zustände der PTZ-Kameras vor:

$$A(t) = \{a_i(t) | i = 1, 2, \dots, p\} \quad (22)$$

Aus Formel 17 wird somit:

$$f_{gc}(R(t), A(t)) = C^{p*}(t) \quad (23)$$

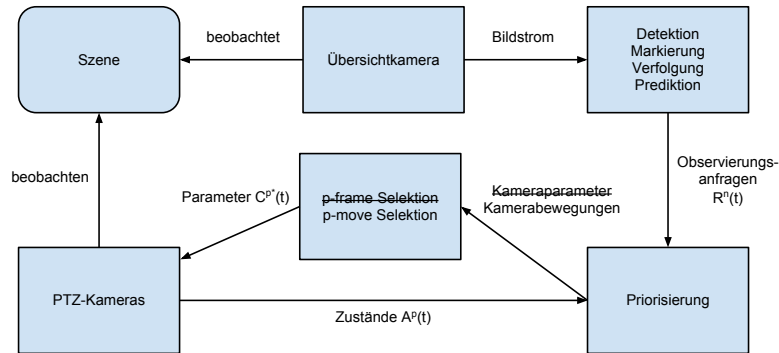


Abbildung 13: Abstrakte Übersicht der Teilkomponenten der Aufmerksamkeitskontrolle.

4.2.1 Verkürzung der Abfangzeit

Bei Xu et al. musste die Aufmerksamkeitssteuerung mit vorhergesagten Personenzuständen arbeiten, die so weit in der Zukunft lagen ($t+t_{lead}$), sodass eine jede PTZ-Kamera in der Lage wäre, noch innerhalb dieser vorgegebenen Zeit t_{lead} von ihrer aktuellen Position an die vorhergesagte Position zu schwenken, und auch den angeforderten Zoom einzustellen. Dieses Zeitintervall bezeichnen wir als Abfangzeit. Da die aktuelle Position und der eingestellte Zoom der Kameras bei [3] nicht bekannt sind, kann dort nur von einem worst-case Wert ausgegangen werden. Mit Hilfe der aktuellen Kamerazustände $A^p(t) = \{a_i(t) | i = 1, 2, \dots, p\}$ ergibt sich die Abfangzeit als:

$$t_{lead} = \max_{c_i \in C^m, a_j(t) \in A^p(t)} \tau(c_i, a_j(t)). \quad (24)$$

Hiermit kann nun die Abfangzeit dynamisch angepasst werden, um so pro Zeiteinheit mehr Bildmaterial von Personen erfassen zu können. Dies lässt die Bildgenerierung durch Feinverstellungen, wie sie in [3] erfolgt und wie auch schon in Kapitel 3.2 beschrieben, zunächst einmal außer acht. Dazu mehr im Abschnitt 4.2.3.

4.2.2 Auswahl der Kamerabewegungen (p-move selection)

Anstatt Kamerakonfigurationen nur anhand der durch sie zu beobachtenden Observierungsanfragen auszuwählen, erscheint es plausibler, auch die zum Umkonfigurieren einer PTZ-Kamera verbrauchte Zeit in Betracht zu ziehen. Schließlich werden in dieser Zeit keine brauchbaren Bilder generiert. Eine Kamerabewegung m_i sei definiert durch:

$$m_i(t) = [a_i(t), c_i, \theta_i(t)]^T \quad (25)$$

a_i ist die betroffene PTZ-Kamera, die die Konfiguration c_i annehmen soll. θ_i stellt die Bewertung dieser Bewegung hinsichtlich der zurückzulegenden zeitlichen Distanz dar.

$$\theta_i(t) = \sigma^{\tau(a_i(t), c_i)} \quad (26)$$

σ ist ein konstanter Wert, der gewichtet, wie groß der Einfluss der temporalen Distanz auf die Bewertung der Bewegung sein soll.

Nun lässt sich eine Sättigungsfunktion definieren, die bei gegebener Observierungsanfrage $r_i(t)$ die Güte der Kamerabewegung bemisst:

$$s(m_i(t), r_j(t)) = s(c_i, r_j(t)) \cdot \theta_i(t). \quad (27)$$

Ähnlich zu den Formeln 13 und 14 lassen sich nun Sättigungsfunktionen für einzelne Kamerabewegungen im Bezug auf mehrere Observierungsanfragen und auch für eine Menge von Kamerabewegungen $M^P(t)$ ableiten:

$$s(m_i(t), R^n(t)) = \sum_{j=1}^n s(c_i, r_j(t)) \cdot \theta_i(t) \quad (28)$$

$$s(M^P(t), R^n(t)) = \sum_{\min}^{\max} \sum_{\min}^{\max} s(c_i, r_j(t)) \cdot \theta_i(t) \quad (29)$$

Wie auch in [3] wird hier ein Maximierungsproblem definiert, dessen Lösung die optimale Auswahl an Kamerabewegungen darstellt:

$$M^{P*}(t) = \arg \max_{M^P} s(M^P(t), R^n(t)) \quad (30)$$

Zusätzlich zu der non-overlapping-condition (NOC) ist hier verständlicherweise noch die Bedingung hinzuzufügen, dass keine zwei Kamerabewegungen in $M^{P^*}(t)$ enthalten sein dürfen, die die selbe PTZ-Kamera ansteuern würden.

Zur Errechnung von $M^{P^*}(t)$ stützen wir uns auf den in [3] vorgestellten Algorithmus (siehe Abbildung 9). Anstelle von Kamerakonfigurationen bewerten wir allerdings Kamerabewegungen und müssen daher in der Zeile 21 die zusätzliche Bedingung stellen, dass zwei die verglichenen Kamerabewegungen nicht die Ansteuerung der selben Kamera anstoßen dürfen. Es wurden weitere Anpassungen vorgenommen, um eine effizientere Berechnung gewährleisten zu können, deren Aufzählung jedoch nicht in den geforderten Rahmen dieser Studienarbeit passt. Erwähnenswert ist, dass die benötigte Rechenzeit im Vergleich zu der für das Verfahren [3] nötigen um den Faktor 10 höher liegt - bei Verwendung von 3 PTZ-Kameras.

4.2.3 Mikroprädiktion und Feinverstellung

In [3] wird die Abfangzeit von den PTZ-Kameras dazu genutzt, die Position anzufahren, in der sich eine Person nach Ablauf der Abfangzeit voraussichtlich aufhalten wird. Anschließend sollen die PTZ-Kameras durch Mikroprädiktionen der Personenpositionen ihre Konfigurationen (auf nicht näher erklärte Art und Weise) feinverstellen, um mehrere Aufnahmen der Personen zu generieren. Die Abfangzeit ist für die Bildgenerierung verlorene Zeit. Da das in dieser Studienarbeit vorgestellte Verfahren allerdings kürzere Kamerabewegungen bevorzugt, deren Durchführung, bei entsprechender Wahl des Gewichtungsparmeters für die temporale Distanz, deutlich weniger Zeit beanspruchen kann, als durch die Abfangzeit eingeräumt wird, könnte man auch vor Erreichen der Abfangzeit mit der Aufnahme der Personen beginnen (siehe Abbildung 14). Hierzu müsste man der PTZ-Kamera lediglich zusätzlich zu der Konfiguration, die bei Ablauf der Abfangzeit erreicht sein muss, auch die Observierungsanfragen auflisten, die durch diese Konfiguration verfolgt werden sollen. Nun könnte die PTZ-Kamera den Mittelwert der Koordinaten der Personen anfahren und den Zoom so wählen, dass alle beobachtbar sind, und sowohl die Koordinaten, als auch den Zoom, immer wieder nach dieser Regel aktualisieren, bis hierdurch ganz automatisch die verlangte Kamerakonfiguration bei Ablauf der Abfangzeit eingestellt ist. Anschließend könnte wie in [3] beschrieben verfahren werden.

4.2.4 Einfluss des Zooms auf die Bewertung von Kamerabewegungen

Wenn sich viele Personen in der Szene befinden, neigt die Kameraparameterfindung dazu, sehr große Bildausschnitte auszuwählen, da die Bewertung des Zooms, den eine Kamerakonfiguration beinhaltet, linear einfließt, die zu erfassende Fläche bei höherem Zoom jedoch quadratisch wächst. Dies führt dazu, dass unter Umständen ein nur sehr kleiner Anteil der gefilmten Fläche sich

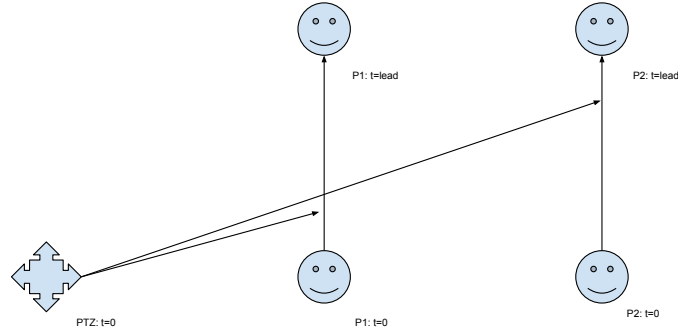


Abbildung 14: Nahe gelegene Observierungsanfragen können länger vor Erreichen der Abfangzeit observiert werden.

als nützlich erweist (siehe Abbildung 15). Um die Kameraressourcen sinnvoller zu nutzen, wäre eine quadratische Gewichtung des Teilterms $\min(1, \frac{z_i}{z_j})$ in der Sättigungsfunktion zu Kameraparametern angebracht. Formel 11 wird somit zu:

$$s(c, r_i(t)) = \omega_i(t) \cdot I(c, r_i(t)) \cdot \min\left(\frac{z_i}{z}, 1\right). \quad (31)$$

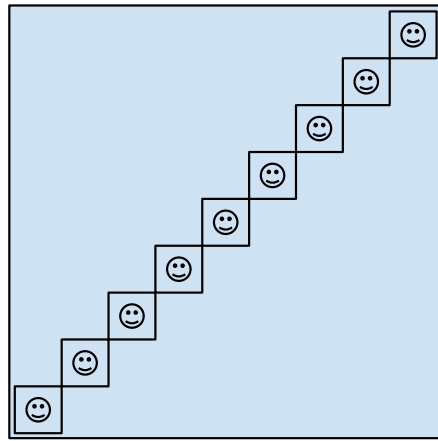


Abbildung 15: Veranschaulichung eines worst-case Szenarios bzgl. der genutzten Bildfläche bei steigender Personen-Anzahl und höherem Zoom.

5 Auswertung

Das in dieser Arbeit vorgestellte Verfahren wurde darauf ausgerichtet um u.a. mit dem Kamerasystem aus Kapitel 3.1 zusammen zu arbeiten. Eine Aufzeichnung der kameraspezifischen Eigenschaften, wie der Zeit, die zur Bewegung der Kamera von einer Position zu einer andern benötigt wird und die Zeit um von einer Zoomstufe zu einer andern zu zoomen wurde durchgeführt. Weil aber bis zum Ende dieser Arbeit noch kein Personen- und Objektdetektionsmodul in das bestehende System integriert wurde, musste die Auswertung dieses Verfahrens auf künstlichen Daten durchgeführt werden, was in folgendem Abschnitt auch folgt.

5.1 Simulation

Um die entwickelten Verfahren evaluieren zu können wurde eine Softwarekomponente implementiert, die bewegte Personen emuliert. Diese erscheinen in Zeitintervallen von 2,5 Sekunden an zufälliger Stelle am Bildrand der Übersichtskamera und bewegen sich in einem Winkel von mind. 40° vom Rand des Bildes weg. Ihre Geschwindigkeit beträgt konstant etwa $10px$ pro 2,5 Sekunden. Der Bildbereich der Übersichtskamera ist insgesamt 900×500 Pixel groß, während eine simulierte Person ca. $80px$ breit und $100px$ hoch ist. Der Parameter zur Gewichtung der temporalen Distanz bei der Umkonfigurierung der PTZ-Kameras beträgt $\sigma = 0,9$, der Parameter zur Gewichtung der noch verbliebenen Zeit, bevor eine Person den Bildbereich verlässt beträgt $\rho = 0,99$ und die Anzahl der zu einer Person aufgezeichneten Bilder unter Berücksichtigung des Zooms beträgt $n_e = 5$.

Beim wie in [3] beschrieben unverändert implementierten Verfahren wurde eine fixe Abfangzeit von 2,6 Sekunden angenommen. Die Abfangzeit berücksichtigt sowohl bei der unveränderten Variante, als auch bei den nach den Vorgaben dieser Studienarbeit erweiterten Arten, nicht die Rechenzeit, die zur Ausführung der Algorithmen benötigt wird. Alle Berechnungen erfolgten unter Verlangsamung der Echtzeit, sodass die Rechenzeit der Algorithmen, im Verhältnis zu z.B. den Kamerabewegungen, vernachlässigbar kurz ausfiel.

5.2 Statistiken

Im folgenden werden statistische Ergebnisse der getesteten Verfahren präsentiert und qualitativ miteinander verglichen.

Ein Vergleich zwischen dem in [3] beschriebenen Verfahren in unveränderter Form und in dahingehend veränderter Form, dass die Abfangzeit dynamisch wie in Kapitel 4.2.1 beschrieben angepasst wird ergab, dass durch die verkürzte Abfangzeit mehr Bilder von Personen mit höherer Güte pro Zeiteinheit generiert werden konnten (siehe Abbildungen 16 und 17).

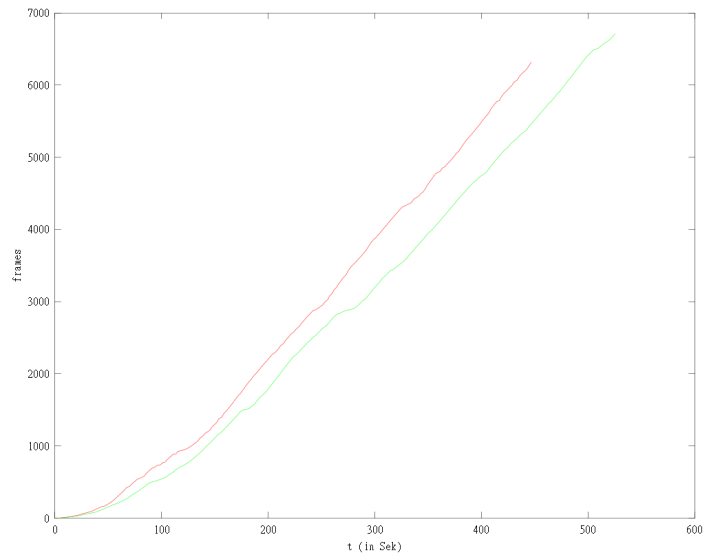


Abbildung 16: Im Vergleich: Xu-Algorithmus mit dynamischer (rot) und fixer Abfangzeit (grün). Auf der x -Achse Zeit t in Sekunden, auf der y -Achse die Anzahl an erfassten Bildern von Personen.

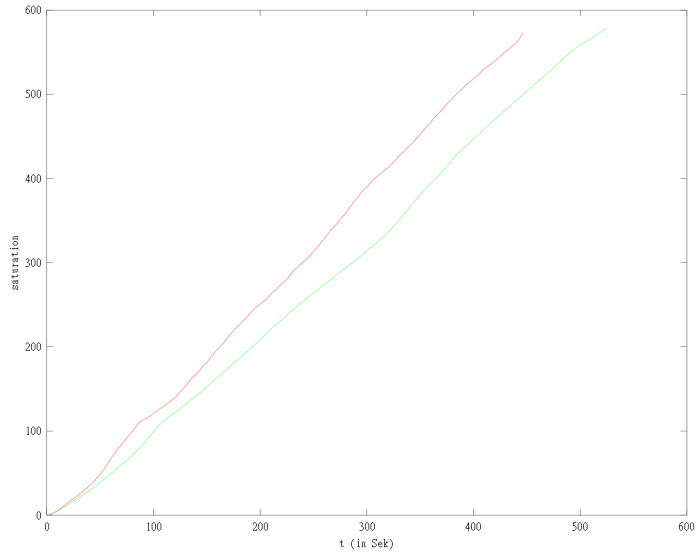


Abbildung 17: Im Vergleich: Xu-Algorithmus mit dynamischer (rot) und fixer Abfangzeit (grün). Auf der x -Achse Zeit t in Sekunden, auf der y -Achse die Güte der erfassten Bilder von Personen.

Außerdem zeigte sich, dass die Berücksichtigung der temporalen Distanz bei der Auswahl der Kamerabewegungen (unabhängig von der Nutzung dynamischer Abfangzeiten) ebenfalls einen Anstieg der Anzahl an beobachteter Personen und des Bildmaterials, in nach Zoomstufe bewerteter Güte, erwirkte. Abbildung 18 und 19 verbildlichen die Ergebnisse (hier: mit dynamischer Abfangzeit).

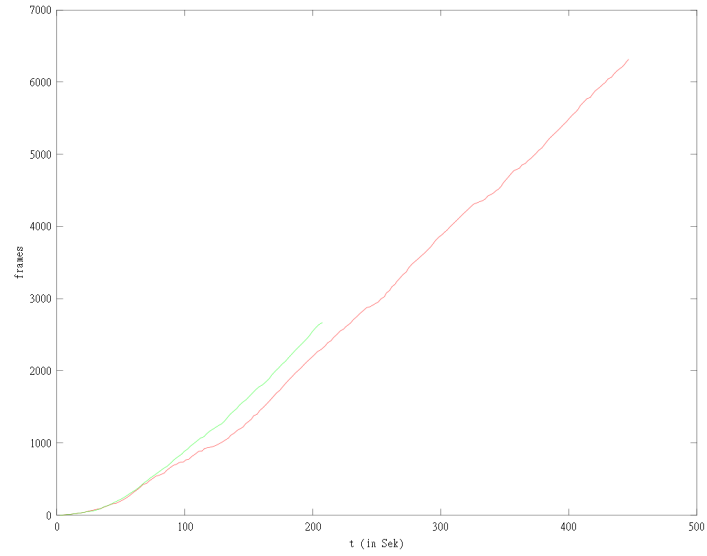


Abbildung 18: Im Vergleich: die Anzahl der generierten Bilder beim Algorithmus von Xu [3] mit (grün) und ohne (rot) Berücksichtigung der temporalen Distanz. Auf der x -Achse die Anzahl an erfassten Bildern von Personen.

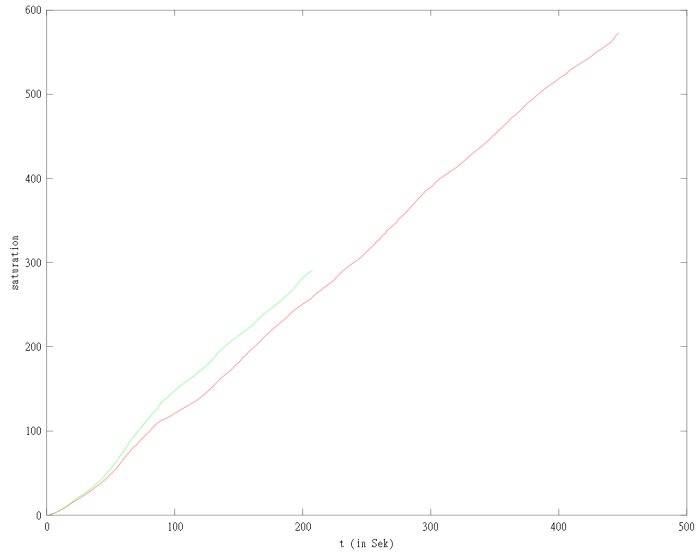


Abbildung 19: Im Vergleich: die Güte der generierten Bilder beim Xu-Algorithmus mit (grün) und ohne (rot) Berücksichtigung der temporalen Distanz. Auf der x -Achse Zeit t in Sekunden, auf der y -Achse die Güte der erfassten Bilder von Personen.

In Abbildung 20 wiederum ist zu erkennen, dass die Berücksichtigung der temporalen Distanz bei der Auswahl der Kamerabewegungen dazu führt, dass die Zeiten, in denen die PTZ-Kameras damit beschäftigt sind, sich zu verstellen, sinkt.

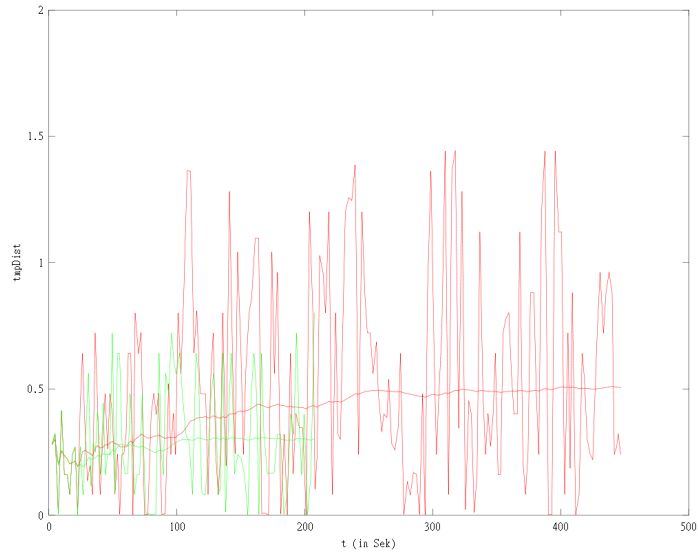


Abbildung 20: Im Vergleich: durchschnittlich zurückzulegende Schwenk- und Zoomzeiten beim Xu-Algorithmus mit (grün) und ohne (rot) Berücksichtigung der temporalen Distanz. Auf der x -Achse Zeit t in Sekunden, auf der y -Achse der Anteil der Bewegungszeit an der Abfangzeit.

Durch Anwendung der in Kapitel 4.2.4 vorgestellten alternativen Zoombewertung steigt die nach dieser neuen Zoombewertung gemessene Güte der Bilder (siehe Abbildung 21), jedoch sinkt die Anzahl an generierten Bildern zu Personen (siehe Abbildung 22).

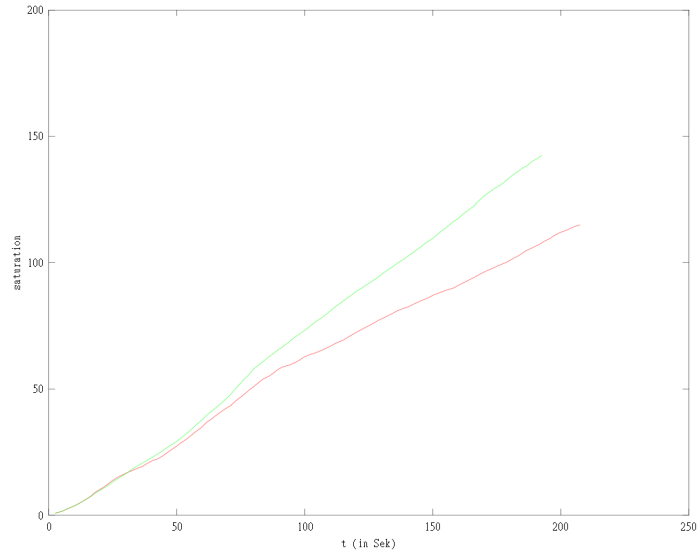


Abbildung 21: Im Vergleich: alternative Zoom-Bewertung (grün) zu in [3] vorgegebener (rot). Auf der x -Achse Zeit t in Sekunden, auf der y -Achse die Güte der Bilder nach alternativer Zoom-Bewertung.

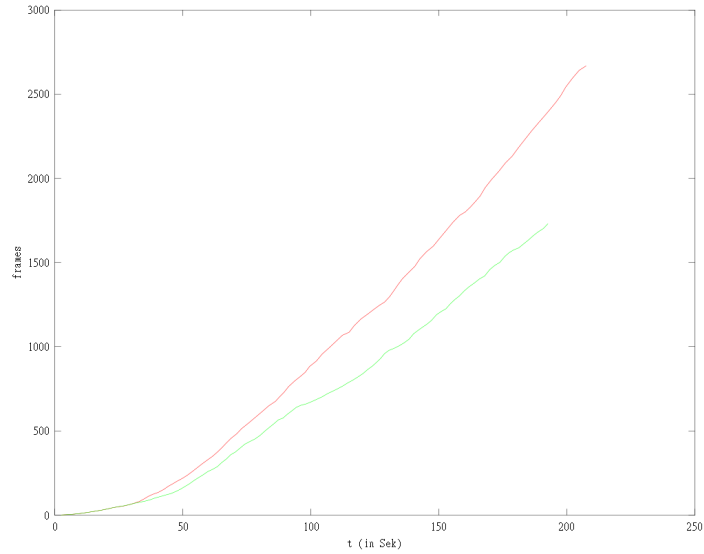


Abbildung 22: Im Vergleich: alternative Zoom-Bewertung (grün) zu in [3] vorgegebener (rot). Auf der x -Achse Zeit t in Sekunden, auf der y -Achse die Anzahl der Bilder nach alternativer Zoom-Bewertung.

5.3 Beispielhafter Auszug aus der Simulation

In den folgenden beiden Abbildungen sind beispielhafte Ausschnitte aus den Experimenten mit simulierten Personen dargestellt.

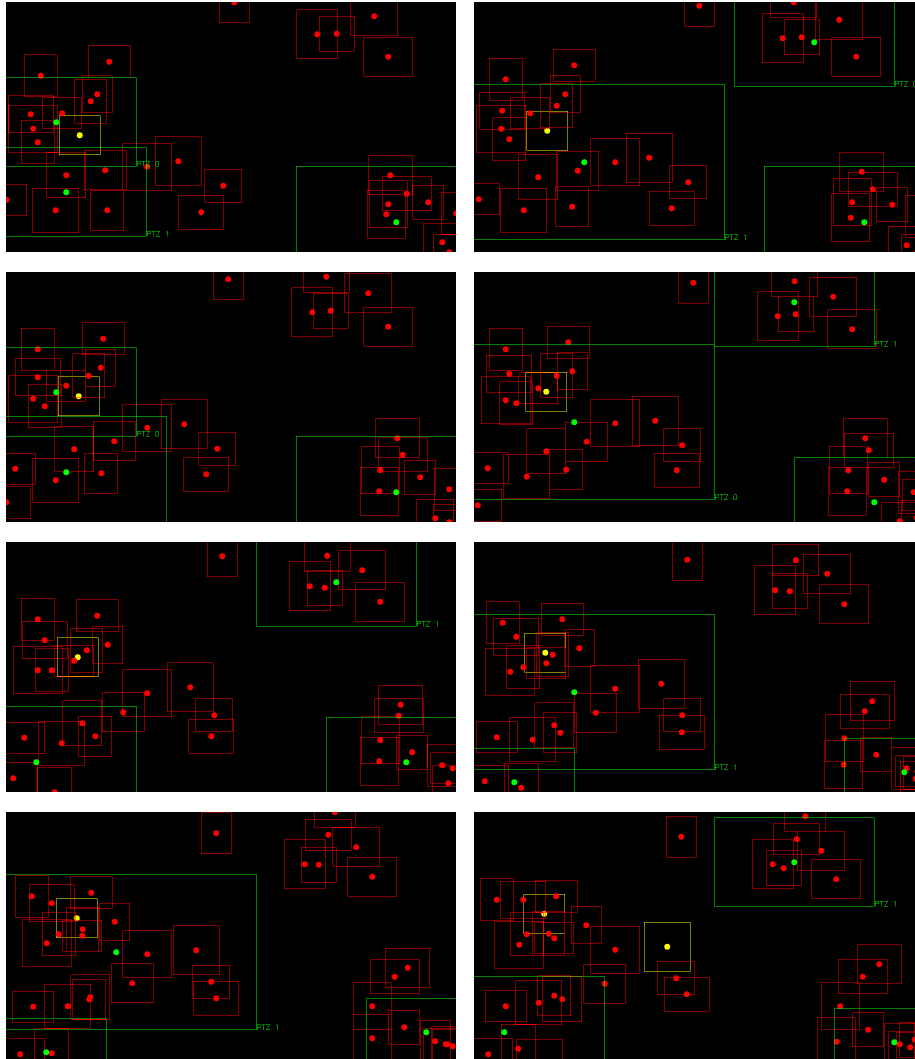


Abbildung 23: Beispielhafter Ausschnitt aus der Simulation. Bildabfolge 1 bis 8 von links nach rechts, von oben nach unten. zu überwachende Personen - rot, ausreichend observierte Personen gelb, PTZ-Bildbereich - grün.

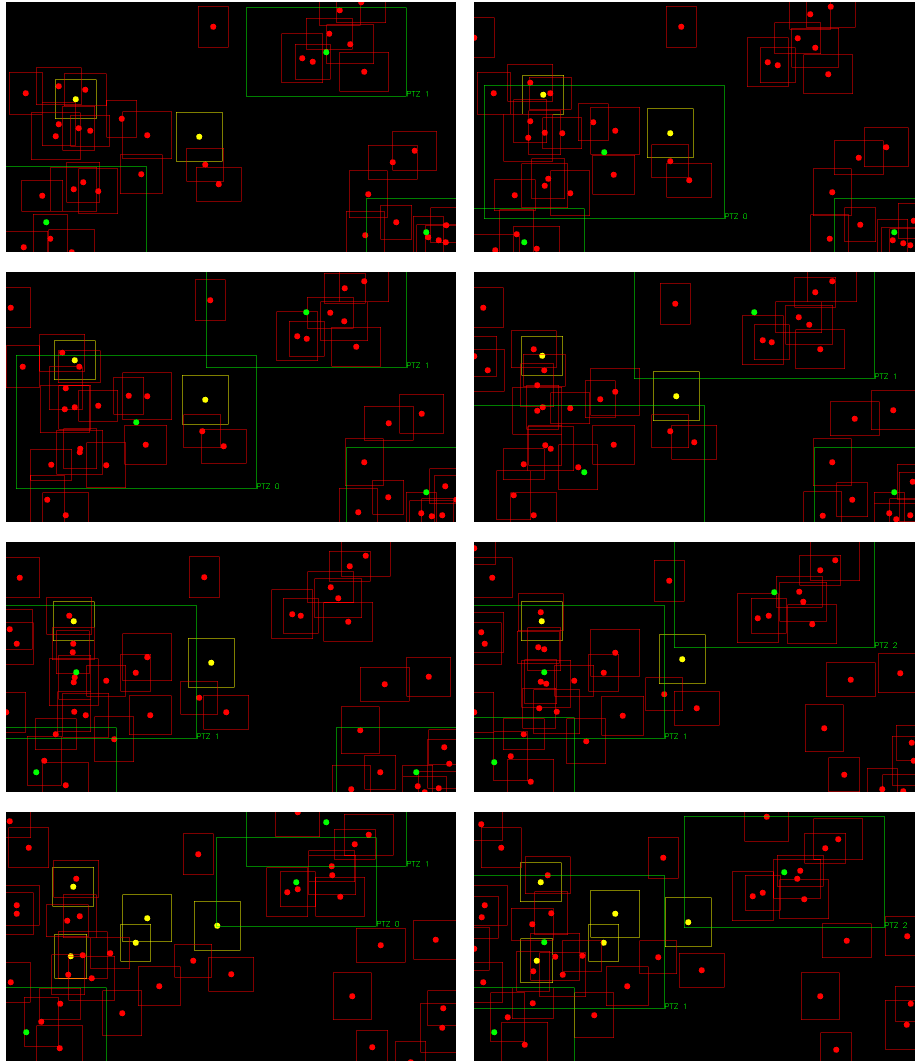


Abbildung 24: Beispielhafter Ausschnitt aus der Simulation. Bildabfolge 9 bis 16 von links nach rechts, von oben nach unten. zu überwachende Personen - rot, ausreichend observierte Personen - gelb, PTZ-Bildbereich - grün.

6 Zusammenfassung

6.1 Ergebnis

Durch eine Reihe von Erweiterungen konnte das in [3] beschriebene Verfahren quantitativ hinsichtlich der Anzahl an erzeugten Bildern von Personen in der Szene und auch die Güte der erzeugten Bilder verbessert werden. Allerdings geht mit den Erweiterungen auch eine erhöhte Rechenzeit der Algorithmen einher, wodurch das hier erarbeitete Verfahren noch nicht echtzeitfähig ist. Weitere Optimierungen am Algorithmus wären allerdings möglich.

6.2 Ausblick

Die Berechnung einer global optimalen Lösung erwies sich als sehr rechenaufwendig. Verfahren, die auf lokaler Ebene operieren, könnten dem Abhilfe schaffen. Im Rahmen eines Multiagentensystems könnten die PTZ-Kameras beispielsweise als lernfähige selbstständige Agenten auf lokaler Ebene in gegenseitiger Kooperation agieren und so den Gewinn in der Summe maximieren, indem sie ihre eigenen Strategien an die äußeren Umstände anpassen.

Dieses Vorgehen würde auch dem manuellen Adjustieren der vielen Parameter Abhilfe schaffen, die die hier erarbeiteten Verfahren voraussetzen.

A Abbildungsverzeichnis

Abbildungsverzeichnis

1	Übersichtsbild Motivation	1
2	Ansatz von Bimbo und Pernici	3
3	Ansatz Qureshi et Terzopoulos	4
4	Mehrkamerasystem - Matchmaking	7
5	Zuordnungsprozess von Kamerakoordinaten - Mehrkamerasystem	7
6	Übersicht Xu Verfahren	9
7	Scheduling zeitl. Ablauf	9
8	Übersicht der Szene mit bewegten Personen	12
9	Abstrakte Übersicht Xu Verfahren	14
10	Xu-Algorithmus	15
11	PTZ-Kamera Schwenkgeschwindigkeiten	17
12	Zoomgeschwindigkeiten der PTZ-Kamera	18
13	Abstrakte Übersicht der Aufmerksamkeitskontrolle	19
14	Abfangzeit und Feinverstellung	22
15	Eigenschaften der Bewertung nach Zoom	23
16	Vergleich statische Abfangzeit vs fixe Abfangzeit - Anzahl Bilder	25
17	Vergleich statische Abfangzeit vs fixe Abfangzeit - Güte der Bilder	26
18	Vergleich Berücksichtigung der temporale Distanz - Anzahl Bilder	27
19	Vergleich Berücksichtigung der temporale Distanz - Güte der Bilder	28
20	Vergleich zurückgelegte temporale Distanz	29
21	Vergleich alternative Zoombewertung - Güte der Bilder	30
22	Vergleich alternative Zoombewertung - Anzahl der Bilder	31
23	Beispielhafter Ausschnitt aus der Simulation. Bildabfolge 1 bis 8 von links nach rechts, von oben nach unten. zu überwachende Personenen - rot, ausreichend observierte Personen gelb, PTZ- Bildbereich - grün.	32
24	Beispielhafter Ausschnitt aus der Simulation. Bildabfolge 9 bis 16 von links nach rechts, von oben nach unten. zu überwachende Personenen - rot, ausreichend observierte Personen gelb, PTZ- Bildbereich - grün.	33

B Literaturverzeichnis

Literatur

- [1] A.-K. Grosselfinger, “Automatische Konfiguration eines Mehrkamerasystems zur Verfolgung von Objekten in unstrukturierten Umgebungen,” Diplomarbeit Fakultät für Informatik. Institut für Anthropomatik (IFA). Karlsruhe 2012.
- [2] C. J. Costello, C. P. Diehl, A. Banerjee, and H. Fisher, “Scheduling an active camera to observe people,” in *Proceedings of the ACM 2nd international workshop on Video surveillance & sensor networks*, pp. 39–45, ACM, 2004.
- [3] Y. Xu and D. Song, “Systems and algorithms for autonomous and scalable crowd surveillance using robotic ptz cameras assisted by a wide-angle camera,” *Autonomous Robots*, vol. 29, no. 1, pp. 53–66, 2010.
- [4] U. M. Erdem and S. Sclaroff, “Look there! predicting where to look for motion in an active camera network,” in *Advanced Video and Signal Based Surveillance, 2005. AVSS 2005. IEEE Conference on*, pp. 105–110, IEEE, 2005.
- [5] F. Z. Qureshi and D. Terzopoulos, “Planning ahead for ptz camera assignment and handoff,” in *Distributed Smart Cameras, 2009. ICDCS 2009. Third ACM/IEEE International Conference on*, pp. 1–8, IEEE, 2009.
- [6] C. J. Costello and I.-J. Wang, “Surveillance camera coordination through distributed scheduling,” in *Decision and Control, 2005 and 2005 European Control Conference. CDC-ECC’05. 44th IEEE Conference on*, pp. 1485–1490, IEEE, 2005.
- [7] A. D. Bimbo and F. Pernici, “Towards on-line saccade planning for high-resolution image sensing,” *Pattern Recognition Letters*, vol. 27, no. 15, pp. 1826–1834, 2006.
- [8] Y. Xu, D. Song, and J. Yi, “Exact algorithms for non-overlapping 2-frame problem with non-partial coverage for networked robotic cameras,” in *Automation Science and Engineering (CASE), 2010 IEEE Conference on*, pp. 503–508, IEEE, 2010.