

Personenerkennung in Videos für die automatische Validierung von Evakuierungssimulationen

AW1 –Vortrag von Torben Woggan
08.12.2011

Gliederung

- ▶ **Einleitung**
 - Mikroskopische Evakuierungssimulationen
 - WALK-Projekt
 - (Bisherige) Idee für die Masterarbeit
 - Ziel für AW1
- ▶ **Personenerkennung in Videos**
 - Anwendungsbereiche
 - Anforderungen an die Personenerkennung für meine Idee
 - Probleme bei der Erkennung von Personen
 - Generelles Vorgehen zur Objekterkennung
 - Ansätze zur besseren Erkennung bei (teilweisen) Verdeckungen
 - Vorstellung einiger Ansätze
- ▶ **Bewertung des Forschungsstandes**
 - Risiken
 - Zuverlässigkeit von Personenerkennung
 - Fazit
- ▶ **Konferenzen & Literaturliste**

Mikroskopische Evakuierungssimulationen

- ▶ Dienen der Simulation von Evakuierungen (z.B. öffentliche Plätze und Gebäude)
- ▶ Ermöglichen Schätzung von Evakuierungszeiten, Gefahrstellen, Verletzten, Verhalten...
- ▶ Besonderheit „Mikroskopischer“-Ansatz:
 - Jede Person dargestellt durch einen Software-Agenten (= autonomer Programmteil mit eigener Intelligenz, Verhalten, Entscheidungen)
 - Simulation (z.B. Verhalten der Agenten) lässt sich 1:1 mit echten Evakuierungen vergleichen
 - Gegenteil: „Makroskopischer“-Ansatz – Abstrahierung durch Fluss-Simulation (Menschenmenge wie Flüssigkeit statt Individuen, Schätzung von Zeiten und Staustellen)

WALK-Projekt

- ▶ Projektleitung:
 - Prof. Dr. Thiel-Clemen, Prof. Dr. Sarstedt
- ▶ Entwicklung einer mikroskopischen Evakuierungssimulation
- ▶ Besonderheiten:
 - Berücksichtigung von Emotionen im Agentenverhalten
 - Möglichkeit des Auftretens von dynamischen Ereignissen (Feuer, Rauch, Wassereinbruch)
- Mein Bereich: Jede Simulation muss validiert werden
 - Stellt die Simulation wirklich den gewünschten Aspekt des simulierten Systems dar?
 - Gleiches Verhalten auf mikroskopischer Ebene (individuelles Verhalten)
 - Gleiches Verhalten auf makroskopischer Ebene (Zeiten, Dichten, Ereignisse)

(Bisherige) Idee der Masterarbeit

- ▶ Bachelorarbeit:
 - Erstellung eines Konzeptes aus Kombination von Verfahren zur Validierung von Gruppensimulationen
- ▶ Idee für die Masterarbeit:
 - Erstellung eines automatischen Verfahrens zur Validierung von mikroskopischen Evakuierungssimulationen mit Hilfe von Personenerkennung in Videoaufnahmen (Aufnahmen von Überwachungskameras von Evakuierungen)
 - Anwendung des Verfahrens für WALK
- ▶ Vorteile der Idee:
 - Objektiv: Vergleich anhand von berechneten Werten (z.B. durchschnittlichen Ausrichtungen, Geschwindigkeiten, Positionen) statt subjektive Meinungen
 - Automatisches Verfahren ließe sich z.B. in jeder Nacht durchführen lassen
 - Nightly Build
 - Personen müssten nicht aufwendig „von Hand“ in jedem Einzelbild gesucht werden
- ▶ Nicht Teil der Masterarbeit:
 - Entwicklung von besseren Verfahren zur Personenerkennung (technische Informatik), sondern Benutzung von Personenerkennung als Werkzeug für die Validierung (angewandte Informatik)

Ziel für AW1

- ▶ Einarbeitung in die Grundlagen der Personenerkennung
- ▶ Speziell: Personenerkennung in Situationen mit leichten bis starken Verdeckungen
 - → z.B. Evakuierungen
- ▶ Klärung der Machbarkeit der Idee und möglicher Probleme

Gliederung

- ▶ Einleitung
 - Mikroskopische Evakuierungssimulationen
 - WALK-Projekt
 - (Bisherige) Idee für die Masterarbeit
 - Ziel für AW1
- ▶ Personenerkennung in Videos
 - Anwendungsbereiche
 - Anforderungen an die Personenerkennung für meine Idee
 - Probleme bei der Erkennung von Personen
 - Generelles Vorgehen zur Objekterkennung
 - Ansätze zur besseren Erkennung bei (teilweisen) Verdeckungen
 - Vorstellung einiger Ansätze
- ▶ Bewertung des Forschungsstandes
 - Risiken
 - Zuverlässigkeit von Personenerkennung
 - Fazit
- ▶ Konferenzen & Literaturliste

Anwendungsbereiche

- ▶ Bisherige Anwendungsbereiche sind u.a.:
 - Unterhaltung
 - Z.B. Xbox 360 mit Kinect
 - Überwachung
 - Z.B. Überwachung von gesperrten Bereichen
 - Fahrzeugsicherheit
 - Z.B. Warnung bei Erkennung von Fußgängern

Anforderungen an die Personenerkennung für meine Idee

▶ Minimale Funktionalität:

- Personen werden erkannt
- Positionen von Personen sind bestimmbar

→ Ermöglicht einfache Validierung auf Makroebene

→ Vergleich von Anzahl an Agenten in bestimmten Bereichen

▶ Optionale Funktionalität:

- Personen werden auch wiedererkannt → Geschwindigkeit von Personen bestimmbar

- Ausrichtung von Personen bestimmbar

→ Ermöglicht genauere Validierung auf Makroebene

→ Vergleich von durchschnittlichen Ausrichtungen, Geschwindigkeiten, Positionen, ...

Anforderungen an die Personenerkennung für meine Idee

- ▶ Weitere wichtige Anforderungen:
 - Hohe Erkennungsraten und niedrige Fehlalarme auch bei Verdeckungen (leicht bis schwer)
 - Validierung soll möglichst genau sein
- ▶ Unwichtig:
 - Hohe Geschwindigkeit der Bestimmung
 - Personenerkennung muss nicht in Echtzeit erfolgen

Anforderungen für die Personenerkennung für meine Idee

- ▶ Es stehen nur 2D-Aufnahmen von Überwachungskameras zur Verfügung
 - Viele moderne Verfahren setzen auf Tiefeninformationen (z.B. Stereo-Kameras)
- ▶ Die Qualität der Aufnahmen kann schlecht sein
 - Z.B. niedrige Auflösung und schlechte Kontraste
- ▶ Auf den Aufnahmen können große Menschenmenge zu sehen sein
 - Viele Verdeckungen möglich
- ▶ Personen laufen möglicherweise
 - Aussehen kann sich stark von normalen Fußgängern unterscheiden

Probleme bei der Personenerkennung

- ▶ Große Unterschiede im Aussehen der Personen durch:
 - Körperhaltung und Rotation
 - Kleidung
 - Rucksäcke & Taschen
 - Kamerawinkel
- ▶ Verdeckungen der Personen:
 - Durch anderen Personen
 - Durch Objekte der Umgebung
 - Durch sich selbst (z.B. Arm vorm Gesicht)

Generelles Vorgehen zur Objekterkennung

- ▶ Trainieren mit positiven und negativen Beispielen des Objektes:
 - Neuronales Netzwerk
 - Support Vector Machine
 - AdaBoost (adaptive Boosting)
- Klassifikator
- ▶ Suchen von Treffern in den Einzelbildern des Videos:
 - Z.B. mit Sliding Window und einem Klassifikator
 - Fenster wird über Bild geschoben und jeweils mit dem Klassifikator entschieden, ob es sich um das Objekt handelt oder nicht

Ansätze zur besseren Personenerkennung bei (teilweisen) Verdeckungen

- ▶ **Generelle Frage:**
 - Wie lässt sich eine Person erkennen, wenn sie nur teilweise sichtbar ist?
- ▶ **Hauptidee:**
 - Unterteilung der Person in kleinere Teile und Erkennen dieser Teile
 - Aus erkannten Körperteilen oder Formen lässt sich auf eine Person schließen, auch wenn sie teilweise verdeckt ist

Körperteil-basierter Ansatz

- ▶ Lernen einzelner explizit definierter Körperteile
 - Klassifikatoren für z.B. Kopf & Schultern, Torso, Arme, Beine, ganzer Körper
 - Bestimmung der räumlichen Verteilung der Körperteile
 - zur Bestimmung des Zentrums des Körpers



Head &
Shoulders



Legs



Torso



Top Left



Top Right



Full Body [3]

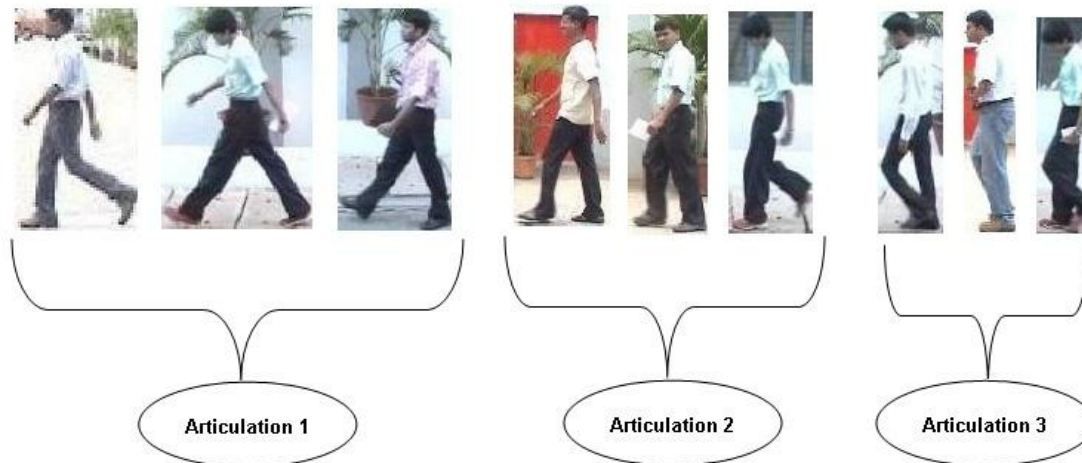
Körperteil-basierter Ansatz

- ▶ Für jedes Körperteil 8 Ausrichtungen lernen



[3]

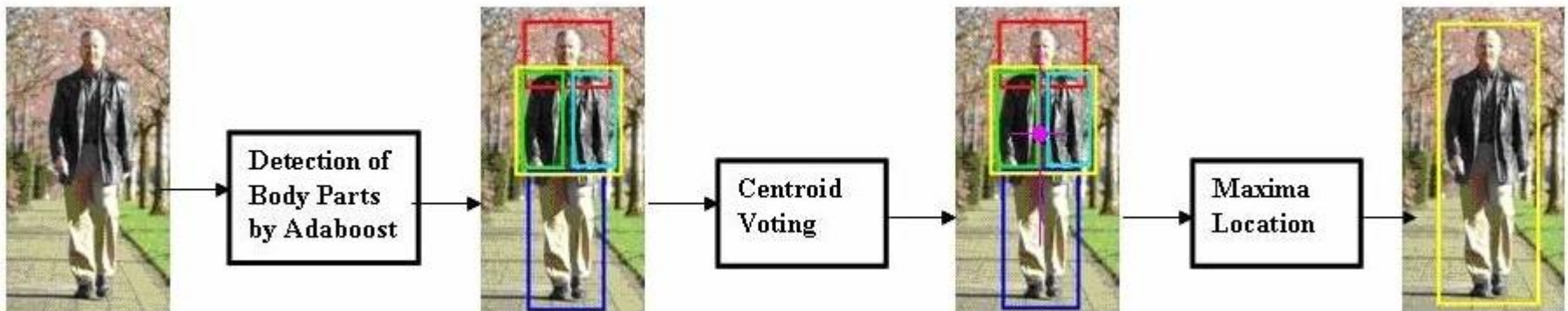
- ▶ Zusätzlich für die Beine 3 Haltungen lernen



[3]

Körperteil-basierter Ansatz

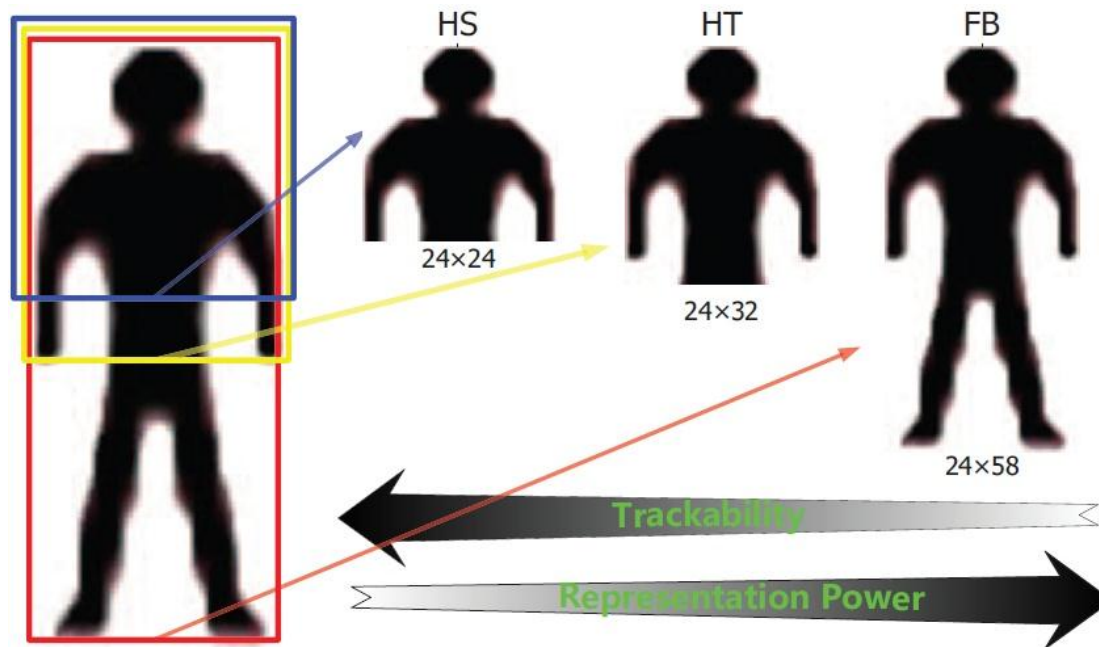
- ▶ Ablauf der Erkennung:
 - Erkennung der Körperteile
 - Anhand bekannter räumlicher Verteilung stimmen Körperteile für das Zentrum des Körpers
 - Größte Anzahl an Stimmen → Zentrum des Körpers
 - Bestimmung der Bounding Box



[3]

Körperteil-basierter Ansatz

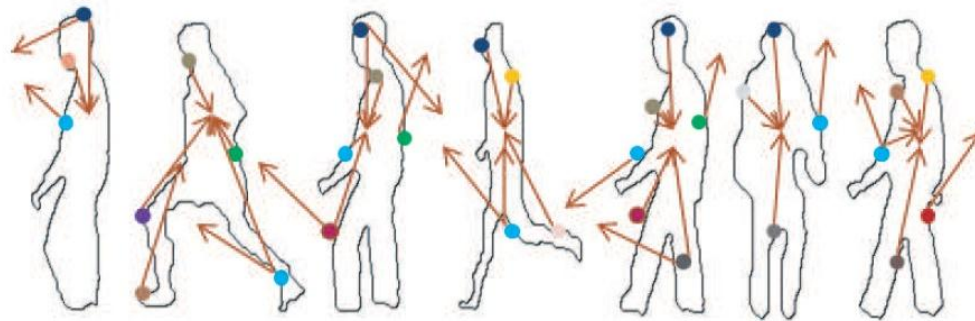
- ▶ Eine weitere Möglichkeit zur Aufteilung
 - Kopf-Schulter
 - Kopf-Torso
 - Ganzer Körper
- ▶ Für jeweils linke, rechte, frontale Ansicht



[7]

Weitere Ansätze

- ▶ Lernen von Körperhaltungen für die Segmentierung und lokalen Verteilungen von Formen im menschlichen Umriss
 - Inkl. Position in Bezug auf das Zentrum des Körpers
 - Inkl. wahrscheinlichsten zur Form passender Umriss des Körpers



[4]

- Bestimmung des Zentrums durch Mehrheitsentscheid

Weitere Ansätze

- ▶ Tracking von „Interest Points“



[6]

- ▶ Ermittlung des Skelett-Graphen (Denaulay Triangulation) und Bestimmung des Kopfes



[2]

Gliederung

- ▶ **Einleitung**
 - Mikroskopische Evakuierungssimulationen
 - WALK-Projekt
 - (Bisherige) Idee für die Masterarbeit
 - Ziel für AW1
- ▶ **Personenerkennung in Videos**
 - Anwendungsbereiche
 - Anforderungen an die Personenerkennung für meine Idee
 - Probleme bei der Erkennung von Personen
 - Generelles Vorgehen zur Objekterkennung
 - Ansätze zur besseren Erkennung bei (teilweisen) Verdeckungen
 - Vorstellung einiger Ansätze
- ▶ **Bewertung des Forschungsstandes**
 - Risiken
 - Zuverlässigkeit von Personenerkennung
 - Fazit
- ▶ **Konferenzen & Literaturliste**

Risiken

- ▶ Zu viele nicht erkannte oder fälschlicherweise erkannte Personen
 - Simulation wird mit zu stark von der Realität abweichenden Daten validiert
 - Falsche Anzahlen von Personen in Bereichen, Dichten
 - Falsche Positionen, Ausrichtungen, Geschwindigkeiten
 - Simulation wird an falsche Daten angepasst
 - Trotz Validierung weicht das Verhalten der Simulation zu stark von der Realität ab
 - Die Validierung ist möglicherweise ungenauer als eine subjektive Bewertung der Simulation
- Das Validierungsverfahren ist nicht brauchbar

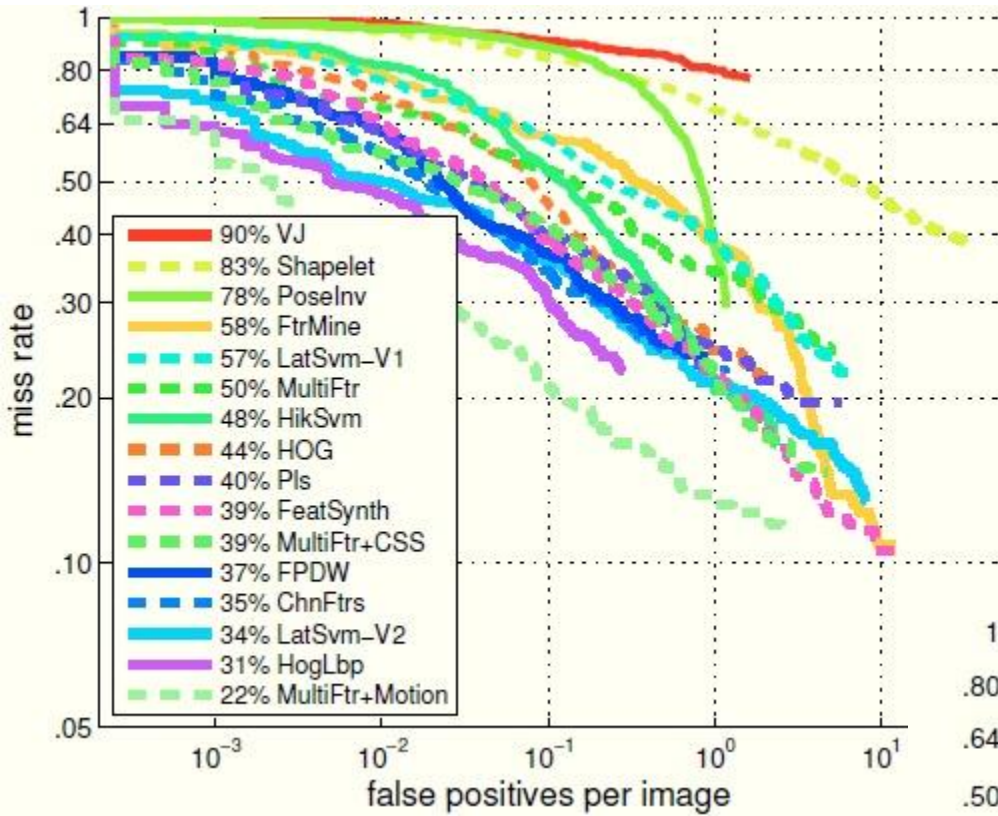
Zuverlässigkeit von Personenerkennung

- ▶ Aktuell (2011): Großer Test zum Stand der Technik
- ▶ Vergleich der Zuverlässigkeit (Erkennungsraten, Fehlerraten) von 16 repräsentativen Verfahren (State-of-the-Art) zur Fußgängererkennung
- ▶ Eigenes Daten-Set (Caltech Pedestrian Dataset):
 - Auto mit Kamera in Bereichen von Los Angeles mit hohen Fußgängerdichten (nachträglich stabilisiert)
 - 250.000 ausgesuchte Einzelbilder (137 Min., 640x480)
 - 2.300 Fußgänger, 350.000 Bounding Boxes
 - 29% nie verdeckt, 53% manchmal verdeckt, 19% immer verdeckt, 70% in min. 1 Frame verdeckt
- ▶ Nicht in den Test mit eingezogen:
 - Bounding Boxes < 20 Pixel Höhe
 - Von den Rändern beschnittene Bounding Boxes
 - Große Gruppen bei denen es nicht möglich war Bounding Boxes für Individuen zu erstellen
 - Wenn selbst durch Menschen nicht eindeutig als Fußgänger identifizierbar

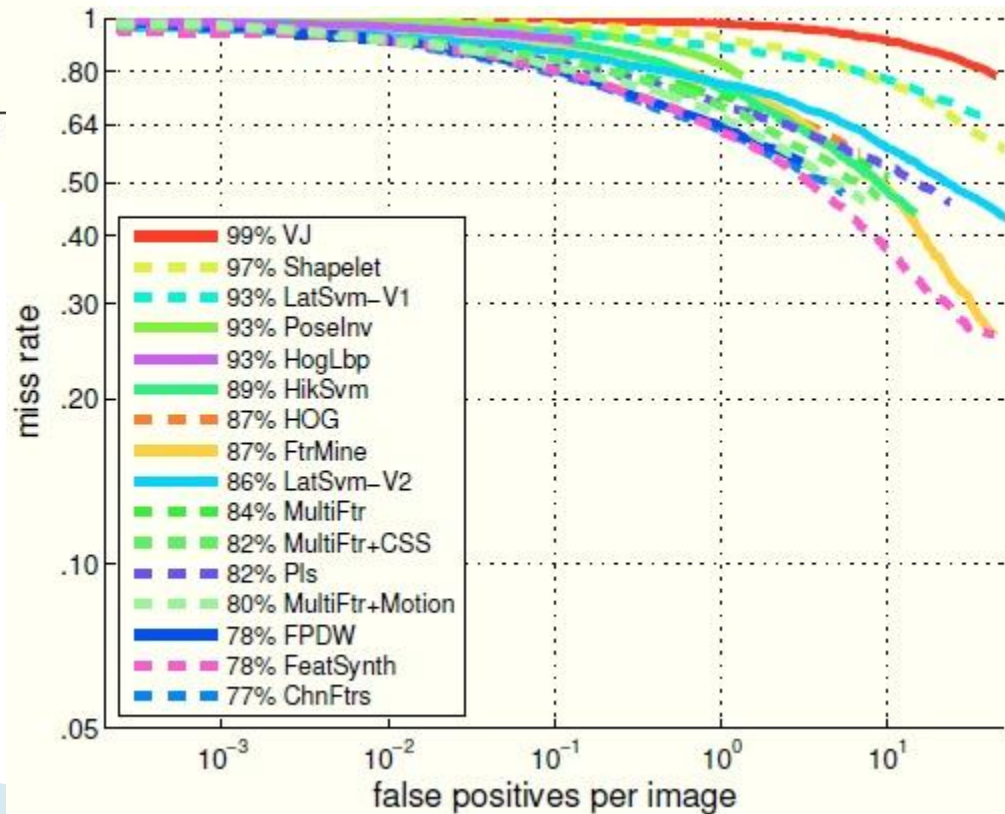
Zuverlässigkeit von Personenerkennung

- ▶ Ergebnisse:
 - Nicht-Erkennungsrate, nur 1 Fehllarm pro 10 Einzelbilder erlaubt
- ▶ Abhängigkeit von der Entfernung (Größe der Fußgänger):
 - Jeweils keine Verdeckungen
 - Selbst bei sehr guten Voraussetzungen sind die Ergebnisse nicht perfekt:
 - Dichte Fußgänger (min. 80 Pixel Höhe)
 - 22% werden vom besten Verfahren nicht erkannt
 - Wenige Erkennungen für mittlere Entfernungen:
 - Mittlere Entfernung (30–80 Pixel Höhe)
 - 77% werden vom besten Verfahren nicht erkannt
 - Sehr wenige Erkennungen für weite Entfernungen:
 - Weite Entfernung (< 30 Pixel Höhe)
 - 95% werden vom besten Verfahren nicht erkannt

Logarithmische Darstellungen



Mittel (30-80 Pixel Höhe) [1]

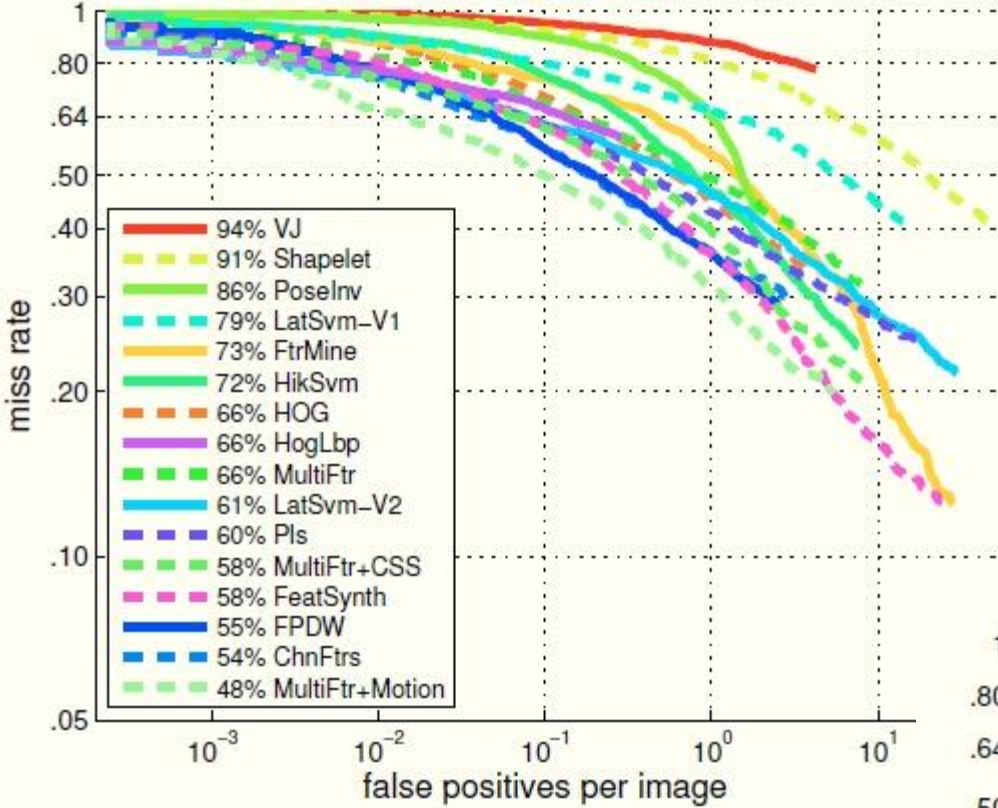


Dicht (min. 80 Pixel Höhe) [1]

Zuverlässigkeit von Personenerkennung

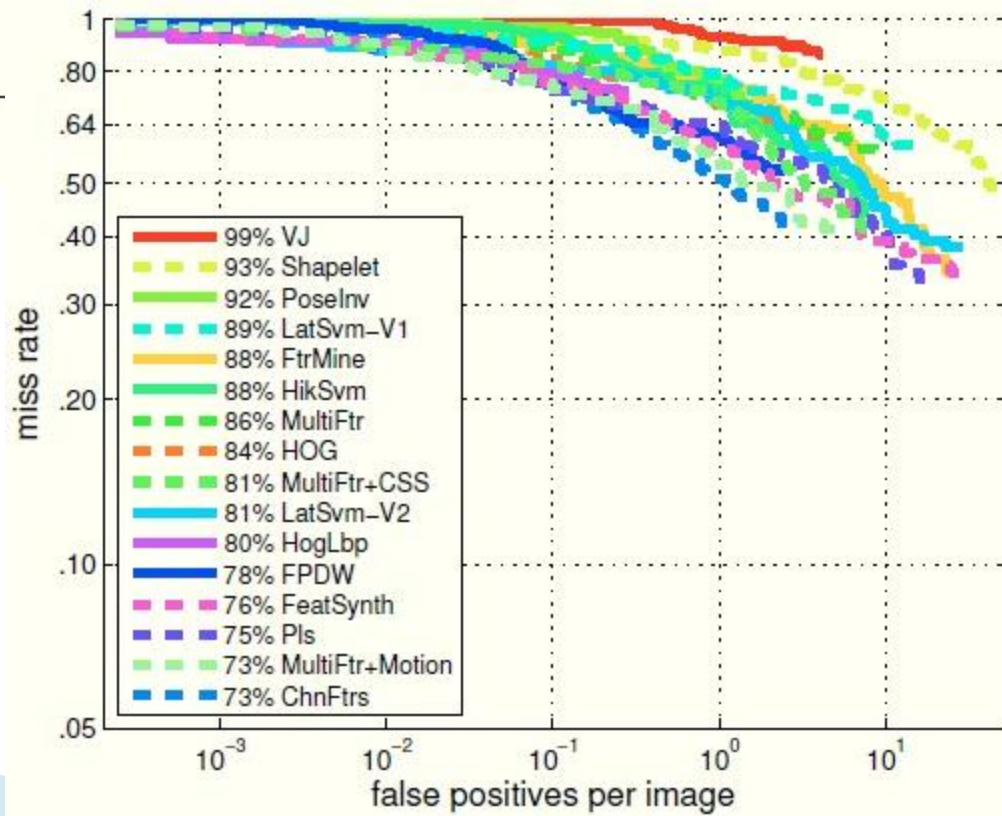
- ▶ Abhängigkeit von der Verdeckung:
 - Mehr als die Hälfte nicht erkannt ohne Verdeckung:
 - Mittlere bis dichte Entfernung (min. 50 Pixel Höhe)
 - Keine Verdeckung
 - 48% werden vom besten Verfahren nicht erkannt
 - Wenige Erkennungen bei teilweiser Verdeckung:
 - Mittlere bis dichte Entfernung (min. 50 Pixel Höhe)
 - 1–35% des Körpers verdeckt
 - 73% werden vom besten Verfahren nicht erkannt
 - Sehr wenige Erkennungen bei starker Verdeckung:
 - Mittlere bis dichte Entfernung (min. 50 Pixel Höhe)
 - 35–80% des Körpers verdeckt
 - 93% werden vom besten Verfahren nicht erkannt

Logarithmische Darstellungen



Keine Verdeckung [1]

1-35% Verdeckung [1]



Fazit

- ▶ Selbst ohne Verdeckungen lassen sich höhere Erkennungsraten nur durch Inkaufnahme von vielen Fehlalarmen erreichen
 - Wie viele Nicht-Erkennungen und Fehlalarme sind hinnehmbar für eine sinnvolle Validierung?
- ▶ Für die Idee sollen Aufnahmen von Überwachungskameras ausgewertet werden:
 - Kleine Personen (bei großen überwachten Bereichen)
 - Verdeckungen sehr wahrscheinlich
- ▶ Ob die Idee mit den heutigen Verfahren umsetzbar ist, ist deshalb fraglich!

Konferenzen

- ▶ **AVSS**
 - IEEE International Conference on Advanced Video and Signal Based Surveillance
- ▶ **CVPR**
 - IEEE Conference on Computer Vision and Pattern Recognition
- ▶ **DICTA**
 - International Conference on Digital Image Computing: Techniques and Applications
- ▶ **ECCV**
 - European Conference on Computer Vision
- ▶ **ICCV**
 - International Conference on Computer Vision
- ▶ **MULTIMEDIA**
 - International Conference on Multimedia
- ▶ Und weitere...

Literaturverzeichnis

- ▶ [1] Dollar, P. ; Wojek, C. ; Schiele, B. ; Perona, P.: Pedestrian Detection: An Evaluation of the State of the Art. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* (2011), Nr. 99, S. 1–20. – Online verfügbar unter: <http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=5975165&isnumber=4359286>
- ▶ [2] Merad, Djamel ; Aziz, Kheir Eddine ; Thome, Nicolas: Fast People Counting Using Head Detection From Skeleton Graph. In: *2010 Seventh IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)*. Boston, MA, USA : IEEE Computer Society, 2010. S. 233–240. – Online verfügbar unter: <http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=5597136&isnumber=5597063>
- ▶ [3] Rao, Supriya ; Pramod, N. C. ; Paturu, Chaitanya Krishna: People detection in image and video data. In: *Proceeding of the 1st ACM workshop on Vision networks for behavior analysis (VNBA '08)*. New York, NY, USA : ACM. S. 85–91. – Online verfügbar unter: <http://doi.acm.org/10.1145/1461893.1461909>
- ▶ [4] Rodriguez, Mikel D. ; Shah, Mubarak: Detecting and segmenting humans in crowded scenes. In: *Proceedings of the 15th international conference on Multimedia (MULTIMEDIA '07)*. New York, NY, USA : ACM, 2007. S. 353–356. – Online verfügbar unter: <http://doi.acm.org/10.1145/1291233.1291310>
- ▶ [5] Seemann, E. ; Fritz, M. ; Schiele, B.: Towards Robust Pedestrian Detection in Crowded Image Sequences. In: *IEEE Conference on Computer Vision and Pattern Recognition, 2007*. Minneapolis, MN, USA : IEEE Computer Society, 2007. S. 1–8. – Online verfügbar unter: <http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=4270325&isnumber=4269956>
- ▶ [6] Sherrah, Jamie: Occluded Pedestrian Tracking Using Body-Part Tracklets. In: *2010 International Conference on Digital Image Computing: Techniques and Applications (DICTA)*. Sydney, NSW, Australien : DICTA 2010 Organisation Committee, 2010. S. 314–319. – Online verfügbar unter: <http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=5692582&isnumber=5692215>
- ▶ [7] Xing, Junliang ; Ai, Haizhou ; Lao, Shihong: Multi-object tracking through occlusions by local tracklets filtering and global tracklets association with detection responses. In: *IEEE Conference on Computer Vision and Pattern Recognition, 2009*. Miami, FL, USA : IEEE Computer Society, 2009. S. 1200–1207. – Online verfügbar unter: <http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=5206745&isnumber=5206488>

Fragen?