Dynamische Kopfposenerkennung anhand statistischer Gesichtsmerkmale

Schriftliche Prüfungsarbeit
für die Bachelor-Prüfung des Studienganges Angewandte Informatik
an der Ruhr-Universität Bochum

vorgelegt von
Christian Behrenberg

24. Februar 2010

Dr. Habil. Rolf P. Würtz
Dr. Susanne Winter

Danksagung

Zunächst danke ich meiner Familie und meinen Freunden dafür, dass sie mir zum einen die nötige Zeit zum Schreiben dieser Arbeit ließen, zum anderen aber hartnäckig genug waren, mich hin und wieder vom Rechner wegzulocken.

Herrn Dr. Rolf P. Würtz bin ich dafür dankbar, dass er mich bei meinem Vorhaben unterstützt hat, ungeachtet dessen, wie unrealistisch meine Ziele zu Beginn waren. Auch wenn der Weg weniger steinig hätte sein können, bin ich froh, dass ich ihn mit all seinen Umwegen gehen durfte.

Ich bedanke mich bei *Thomas Walther* dafür, dass er mir einen Blick über den Tellerrand der Computer Vision gewährt hat - aber auch, dass der Stand der Technik unerwartet viel Raum zur Verbesserung bietet. Durch Berichte über aktuelle Trends, Probleme und wie man diese in der Praxis tatsächlich löst, konnte er mich dazu motivieren, nicht aufzugeben, da ich zwischenzeitlich fest davon überzeugt war, dass ich scheitern würde.

Für den regen Gedankenaustausch und der Rekonstruktion meiner Texturmodelle möchte ich mich bei *Manuel Günther* bedanken.

Inhaltsverzeichnis

Abbildungsverzeichnis

Tabellenverzeichnis

1 Einleitung

1.1 Nonverbale Komponenten sozialer Interaktion

Es ist dem Menschen ohne Anstrengung möglich, seine Umgebung zu erfassen und mit anderen Menschen zu kommunizieren. Neben gesprochener Sprache sind die Kopfpose und die Blickrichtung als nonverbales Kommunikationsmittel eine wichtige Komponente sozialer Interaktion [Eme00]. Mit ihnen kann das Verhalten eines Gegenübers mehrschichtig analysiert werden. Die Interaktion zwischen Mensch und Maschine wird durch Berücksichtigung dieser Informationen gefördert:

- Müdigkeitsattacken sind ein häufiger Grund für Auffahrunfälle [KW94]. Fahrerassistenzsysteme, die die Vigilanz des Fahrers über Registrierung seiner Kopfhaltung überwachen (*driver drowsiness detection*), können in Gefahrensituationen warnen [JY02].

- *Facial Motion Capture*-Verfahren werden zur Erfassung der Bewegung und Mimik des Gesichts einer Person verwendet, um virtuelle Charaktere in Film- und Spieleproduktionen [BL85][Men00] zu animieren.

- Zur automatisierten Erkennung von Individuen an öffentlichen Plätzen werden Gesichtserkennungsverfahren (*face recognition approaches*) verwendet, die durch korrekte Schätzung der Kopfpose verbessert werden können [MKMW07].

- Mithilfe einer blickgesteuerten Augenmaus wird es körperlich beeinträchtigten Menschen ermöglicht, einen Computer zu bedienen [TEHF89].

1.2 Verfahren zur Erkennung der Kopfpose in Bildern

Wenn es darum geht, die Kopfpose und Blickrichtung zu bestimmen, ist es nicht trivial, die kognitive Leistung des Menschen in Form von bildverarbeitenden Algorithmen auf Maschinen zu übertragen. Etablierte Verfahren zur Detektion [YKA02] und -erkennung [ZCPR03] von Gesichtern sind meist nur für Frontalansichten geeignet. Die Bestimmung der Kopfpose (*head pose estimation*) ist ein Sonderfall der Gesichtsdetektion (*face detection*), da für mehrere rotierte Gesichter die Kopfpose erkannt werden muss.

Die meisten existierenden Verfahren zur Schätzung der Kopfpose basieren auf dieser Annahme und unterscheiden sich in Bezug auf die Schätzung von

diskreten/kontinuierlichen Winkeln, Modellen/Exemplaren von Kopfposenansichten oder der bild-/videobasierten Verarbeitung. Sie lassen sich wie folgt kategorisieren [MCT09]:

Appearance template methods (ATM's) liegen mehrere Ansichten rotierter Gesichter als Repräsentationen diskreter Kopfposen zugrunde (*face templates*) [Bey94][NF96]. Jedes *template* wird über das Bild geschoben (*sliding window approach*, [HARK96][CHLH08]), an jeder Stelle mit dem Bildausschnitt über *template matching* [Bru09] verglichen und ein *match score* berechnet. Ein *match score* ist ein skalarer Zahlenwert, der die Ähnlichkeit des *templates* zum Bildausschnitt wiedergibt. Der zugeordnete Winkel desjenigen *templates*, mit dem der größte *match score* für das Bild erzielen werden kann, entspricht der ermittelten Kopfpose.

Detector arrays setzen sich aus mehreren Gesichtsdetektoren zusammen [JV03][HSW98]. Jeder Detektor repräsentiert eine diskrete Kopfpose durch ein gelerntes Texturmodell der Kopfposenansicht. Ähnlich der Kopfposenschätzung mithilfe von ATM's, entspricht der erkannte Posenwinkel demjenigen Detektor, der den größten *match score* erzielen kann.

Nichtlineare Regressionsmethoden werden zur Ermittlung eines funktionalen Zusammenhangs zwischen einem Bildsignal und einem kontinuierlichen Kopfposenwinkel verwendet. Mithilfe einer auf diese Weise gelernten mathematischen Funktion, beispielsweise in Form eines künstlichen neuronalen Netzes [Bis05], werden die Kopfposenwinkel direkt über die Pixelwerte berechnet [SYW02][ESS04].

Geometrische Verfahren berücksichtigen im Gegensatz zu den oben genannten Methoden die Bildkoordinaten der Gesichtsmerkmale (Landmarken). Längenverhältnisse und Winkel, die sich über die relative Lage der Landmarken zueinander ergeben, werden unter Zuhilfenahme zuvor etablierter Beobachtungen bezüglich der Geometrie des Kopfes in kontinuierliche Winkel umgerechnet [GC94].

Trackingverfahren finden im Rahmen der Verarbeitung von Bildsequenzen Anwendung und sind, ähnlich den geometrischen Verfahren, merkmalsbasiert. Sind Landmarken des Gesichts im Bild bekannt, ist es möglich,

deren Verschiebung von Bild zu Bild zu erkennen [TK91]. Ist die initiale Kopfpose auf dem ersten Bild der Bildsequenz bekannt, kann durch Registrierung dieser bildweisen Änderungen die relative Winkeländerungsrate abgeschätzt und durch Addition dieser Werte die Kopfpose ermittelt werden [GC96][YZ02].

1.3 Ziele dieser Arbeit

Das Ziel dieser Arbeit ist die Erkennung der kontinuierlichen Kopfpose einer einzelnen Person in farbigen, hochauflösenden Bildsequenzen. Die Bestimmung der Kopfpose ist an ein Verfahren von McKenna und Gong angelehnt [MG98].

McKenna und Gong verwenden eine ATM, deren *face templates* durch eine angenäherte *Gabor wavelet transformation* (GWT) nach Daugman [Dau85] verarbeitet werden. Zu Beginn einer Sitzung werden mehrere Kopfposenansichten des Benutzers interaktiv[1] aufgenommen und mit ausgewählten *Gabor wavelets* [Gab46] im Ortsraum gefaltet. Die Absolutwertantworten des gefalteten Bildes werden als *face templates* gespeichert und zur Laufzeit mit dem auf dieselbe Weise prozessierten Kamerabild verglichen. Das System ist nicht in der Lage, die Bewegung des Kopfes in der Tiefe zu erkennen. Der zentrale Punkt dieser Arbeit ist die Verwendung statistischer Texturmodelle zur personenunabhängigen Erkennung der Kopfpose. Das Verfahren von McKenna und Gong wird sinnvoll erweitert und wie folgt verbessert:

Normalisierung der Gesichtsregion Um die Bewegung des Kopfes in der Tiefe zu kompensieren, wird die Gesichtsregion zunächst separat lokalisiert, extrahiert und auf eine einheitliche Größe gebracht (Normalisierung). Das Gesicht wird lokalisiert, indem das Farbbild durch Detektion von Hautfarbe (*skin color detection*) [VVA03][KMB07] in mögliche Kopfbereiche zerlegt

(a) *face templates* (b) Gefaltetes Bild und Detektionsergebnis

Abb. 1: Modifizierte Abbildung des Kopfposenerkennungsverfahrens von McKenna und Gong [MG98]. Die Absolutwertantworten der mit einem horizontalen *Gabor wavelet* gefalteten *face templates* ist in Abbildung 1(a) und die von drei Eingangsbildern in Abbildung 1(b) zu sehen.

[1] Den *face templates* liegen keine *ground truth*-Daten zugrunde.

wird. Anschließend wird innerhalb dieser Regionen unter Zuhilfenahme eines modifizierten Viola/Jones-Detektors [VJ01] das Gesicht erfasst.

Statistische Texturmodelle Statt *appearance templates* werden in dieser Arbeit statistische Texturmodelle in Anlehnung an *Maximum-Likelihood* (ML)-Gaborgraphen [GW09] verwendet. Die Texturmodelle sind im Gegensatz zur ursprünglichen Definition nach [LWvdM97] nicht elastisch. Sie werden auf Basis der Pointing´04 Bilderdatenbank [GHC04] (siehe Abschnitt 2.1.2) trainiert, um ein personenübergreifendes Texturmodell zu erhalten; das Verfahren lässt sich deshalb in die Klasse der *detector array*-Methoden einordnen. Die Struktur der Gaborgraphen jedes Texturmodells wird statistisch ermittelt, indem alle Trainingsbilder *a-priori* mit Landmarken versehen werden. Tritt ein Gesichtsmerkmal in einer Pose in der Trainingsmenge nicht durchgängig auf, ist es instabil. Die Auftrittswahrscheinlichkeiten instabiler Landmarken werden ebenfalls pro Pose statistisch ermittelt und in den Texturmodellen berücksichtigt.

Ground-truth Jedes Trainingsbild der Datenbank ist mit einer Angabe des dargestellten Posenwinkels versehen. Auf diese Weise ist die Angabe eines konkreten Winkels möglich. Die Verwendung der Pointing´04 Datenbank ermöglicht die Einordnung und den Vergleich mit anderen *head pose estimation*-Verfahren[2], die auf derselben Datenbank evaluiert wurden (siehe Abschnitt 4.2).

Gabor wavelet transformation McKenna und Gong schränken die GWT auf die Verwendung von wenigen horizontalen und vertikalen *Gabor wavelets* ein, um eine echtzeitfähige Implementierung zu erreichen. Diese Beschränkung wird hier aufgehoben. Es wird die klassische diskrete *Gabor wavelet*-Familie [LWvdM97] verwendet.

Tracking Sowohl die Kopflokalisation als auch der ermittelte Posenwinkel können fehlerhaft sein. Da Bildsequenzen verarbeitet werden, soll die ermittelte Kopfregion und -pose mit einem Kalman-Filter [Kal60][BW01] unter Zuhilfenahme eines dynamischen Bewegungsmodell verfolgt werden.

[2] [Sti04], [GMHC07], [JTH07], [VNS07] und [LFY+07]

2 Material & Methoden

2.1 Datenbanken

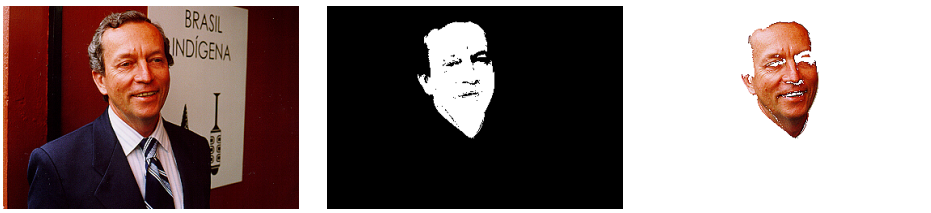
2.1.1 Compaq Bilderdatenbank

Die Compaq-Bilderdatenbank [JR02] enthält 13.640 Farbbilder $\boldsymbol{\mathcal{I}}$ unterschiedlicher Größe. Sie sind dem Internet entnommen und in zwei Bildreihen $\boldsymbol{\mathcal{I}}_{\{S;\neg S\}}$ aufgeteilt. $\boldsymbol{\mathcal{I}}_S$ enthält $|\boldsymbol{\mathcal{I}}_S| = 4.675$ Bilder mit sichtbarer Haut. Hautmasken $\boldsymbol{\mathcal{M}}$ in Form von schwarz/weiß-Bildern repräsentieren die Klassenzugehörigkeit jedes Pixels $\vec{p} \in \boldsymbol{\mathcal{I}}_S$ zur *skin color class* S (siehe Abbildung 2). Die Farbe eines Pixels $\vec{p}$ ist als dreielementiger Vektor $\vec{c}$ beschreibbar. Die Aufzählung der skalaren Werte von $\vec{c}$ nennt man Farbtupel. Dieses erlaubt die numerische Charakterisierung einer Farbe in einem Farbraum[3]. Trotz manueller Erstellung sind die Hautmasken nicht vollständig. Aus diesem Grund dürfen Farbtupel nicht maskierter Pixel aus $\boldsymbol{\mathcal{I}}_S$ *nicht* als Beispiele für *non-skin color* verwendet werden. Die Farbtupel der Bildreihe $\boldsymbol{\mathcal{I}}_{\neg S}$ repräsentieren alle *non-skin color*.

Es existieren für $\Omega = \{S, \neg S\}$ insgesamt $|S| = 80.377.671$ *skin color*-Tupel und $|\neg S| = 854.744.181$ *non-skin color*-Tupel. Die Masken liegen als PBM-Dateien (*portable bitmap file format*) vor, während die Bilder als JPG, GIF oder PNG Dateien gespeichert sind.

2.1.2 Pointing ´04 Bilderdatenbank

Die Pointing´04 Bilderdatenbank [GHC04] enthält 30 Bildsequenzen mit Aufnahmen unterschiedlicher Kopfposen von 15 Subjekten. Jede Bildsequenz besteht aus 93 Aufnahmen, die jeweils eine Kopfpose repräsentieren (siehe Abbildung 4). Die Nick- (Θ) und Gierwinkel[4] (Ψ) sind diskretisiert mit $\Theta \in \{0°,$



(a) Farbbild (b) Hautmaske (c) maskierte Hautpixel

Abb. 2: Bild der *skin color class* aus der Compaq-Bilderdatenbank.

[3]Die Farbe „Rot“ ist beispielsweise im RGB-Farbraum mit dem Farbtupel $\vec{c} = (R, G, B)^T = (255, 0, 0)^T$ gekennzeichnet.

[4]Nickwinkel: Auf-/Abwärtsrotation, Gierwinkel: Links/Rechts-Rotation des Kopfes.

$\pm 15°, \pm 30°, \pm 60°, \pm 90°\}$ und $\Psi \in \{0°, \pm 15°, \pm 30°, \pm 45°, \pm 60°, \pm 75°, \pm 90°\}$. Die Aufnahmen decken alle Kombinationen von Nick- und Gierwinkeln ab, bis auf Posen mit $\Theta = \pm 90°$, für die es nur eine Aufnahme mit $\Psi = 0°$ gibt. Die Winkelangaben pro Aufnahme sind *indirekte* Messwerte, da die Subjekte pro Pose ihren Kopf in eine vorgegebene Blickrichtung wenden sollten, ohne die Augen zu bewegen.

Der Hintergrund ist in allen Bildern hellgrau. Die Subjekte zeigen unterschiedliche Gesichtsausdrücke und unterscheiden sich in Alter, Geschlecht, Hautfarbe und Gesichtsstruktur. Jedes Bild ist 384x288 Pixel groß und enthält einen linksseitigen schwarzen Balken von 5 Pixeln Breite. Die Bilder liegen im JPG Format vor, die Angaben der Kopfposenwinkel sind in den Dateinamen kodiert.

2.1.3 Daten anthropomorpher Gesichtsmessungen

In der Studie von Joseph W. Young [You93] werden die statistischen Ergebnisse einer Erhebung über 30 Jahre präsentiert, in denen anthropomorphe Messdaten der Köpfe und Gesichter erwachsener US-Bürger erhoben wurden. Es existieren 22 Messdaten in Bezug auf 27 Gesichtsmerkmale. Insgesamt sind maximal 376 Subjekte (195 Frauen und 172 Männer) im Alter von 17-69 Jahren beteiligt. Der überwiegende Teil der Versuchsteilnehmer weist eine helle Haut und europäische, kaukasische Physiologie auf.

In der Studie werden der statistische Mittelwert, die Standardabweichung, die relative Standardabweichung, der mittlere Fehler, der Maximal-, bzw. Minimalwert sowie andere statistische Angaben nach Geschlechtern getrennt präsentiert. Alle Messdaten sind in *inch* angegeben und teilweise fehlerhaft in cm und mm angegeben. Zwei der in dieser Arbeit verwendeten Datenblätter sind in Anhang B aufgeführt.

2.1.4 Annotation der Pointing ´04 Datenbank

Jede Kopfposenansicht der Pointing ´04 Bilderdatenbank (siehe Abschnitt 2.1.2) wurde mit Landmarken $\vec{l}_l$ versehen. Die Landmarken $\boldsymbol{\mathcal{L}}$ decken pro Bild unterschiedlich viele Gesichtsmerkmale l ab. Die Menge aller möglichen Gesichtsmerkmale bildet das Gesichtsmodell $\boldsymbol{\mathcal{F}}$. Für jedes Bild gilt $\forall \vec{l}_l \in \boldsymbol{\mathcal{L}} : l \in \boldsymbol{\mathcal{F}}$. In Abgrenzung zu proprietären Gesichts-, bzw. Posengraphen[5] basiert die Auswahl der Gesichtsmerkmale in $\boldsymbol{\mathcal{F}}$ auf folgenden Kriterien:

[5] Vergleiche [LWvdM97, Abbildung 4], [NKvdM97, Abbildung 1], [GW09, Abbildung 3] und [MW09, Abbildung 1]

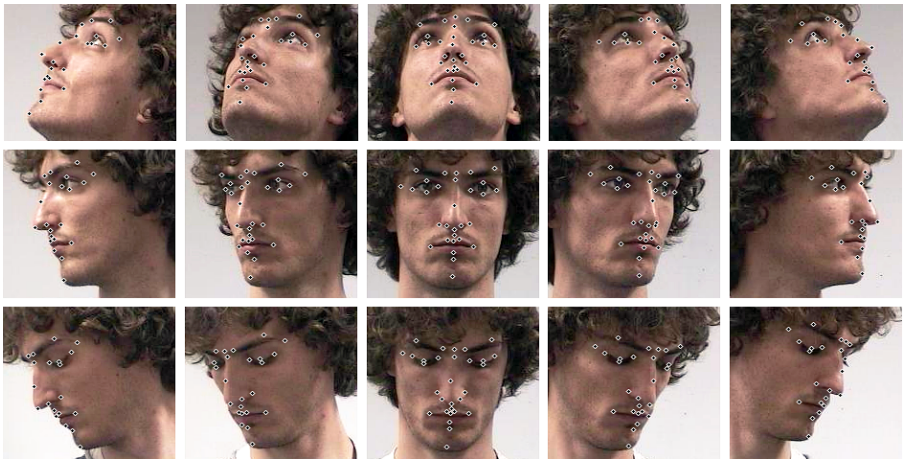


Abb. 3: Modifizierter Auszug einer Bildsequenz der Pointing´04 Datenbank [GHC04]. Die Kreise markieren die annotierten Landmarken $\vec{l}_l \in \boldsymbol{\mathcal{L}} \subseteq \boldsymbol{\mathcal{F}}$ (siehe Abschnitt 2.1.4). Abgebildet sind die Kopfposen $\Theta = \{0°, \pm 45°\}, \Psi = \{0, \pm 45°, \pm 90°\}$.

1. Jedes Gesichtsmerkmal soll mithilfe texturbasierter Filter deutlich erkenn- und lokalisierbar sein.
2. In der Frontalansicht müssen alle Gesichtsmerkmale zu sehen sein.
3. In Vollprofilansichten soll mindestens die Hälfte aller Gesichtsmerkmale sichtbar sein.
4. Gesichtsmerkmale sollen robust gegenüber typischen Selbstverdeckungen (Selbstokklusion) und Veränderungen der Haut sein.
5. Gesichtsmerkmale müssen physiologisch lokalisierbar sein. Bildbezogene Landmarken, wie z.B. die Silhouette, sind nicht erlaubt.

$\boldsymbol{\mathcal{F}}$ besteht aus anthropomorphen Gesichtsmerkmalen [Far94] (siehe Tabelle 1). Sie berücksichtigen die physiologischen Gegebenheiten der Augen-, Nasen- und Mundregion des menschlichen Gesichts (vergleiche Abbildung 4). $\boldsymbol{\mathcal{F}}$ korrespondiert mit Erkenntnissen der Gesichtsanthropometrie [Far94] und Empfehlungen zur Patientenuntersuchung in der modernen Gesichtschirurgie [Men05].

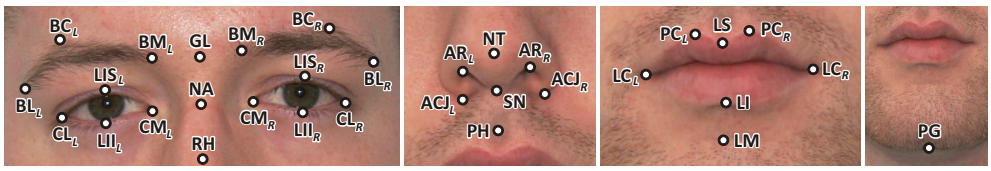


Abb. 4: Exemplarische Landmarken der Gesichtsmerkmale in $\boldsymbol{\mathcal{F}}$ in Ausschnitten einer Frontalansicht.

Augenregion		Nasenregion		Mundregion	
$BL_{l;r}$	Lateral Brow	GL	Glabella	PH	Philtrum
$BC_{l;r}$	Central Brow	NA	Nasion	$PC_{l;r}$	Philtrum Column
$BM_{l;r}$	Medial Brow	RH	Rhinion	LS	Labrale Superior
$CL_{l;r}$	Lateral Canthus	NT	Nasal Tip	$LC_{l;r}$	Lip Commisure
$CM_{l;r}$	Medial Canthus	$AR_{l;r}$	Alar Rim	LI	Labrale Inferior
$LIS_{l;r}$	Superior Lid	$ACJ_{l;r}$	Alar Crease Junction	LM	Labiomental
$LII_{l;r}$	Inferior Lid	SN	Subnasale	PG	Pogonion

Tab. 1: Gesichtsmerkmale des Gesichtsmodells $\boldsymbol{\mathcal{F}}$. Die Merkmale sind der Augen-, Nasen- und Mundregion zuzuordnen. Angegeben sind jeweils der Name und das Kürzel jedes Gesichtsmerkmals. Ist ein Gesichtsmerkmal symmetrisch, gibt der Index l, bzw. r die Zuordnung zu einer der Gesichtshälften aus der Sicht des Betrachters an.

Die Verwendung in anthropometrischen Studien [You93] (vergleiche Abschnitt 2.1.3) bekräftigen die Auswahl der Landmarken.

2.2 Implementation und Hardware

Das hier vorgestellte Verfahren zur Kopfposenerkennung wird in einem Computerprogramm zur Bestimmung der kontinuierlichen Kopfpose in Bildsequenzen umgesetzt. Es ist für die Verarbeitung von aufgezeichnetem Videomaterial im RGB-Farbformat für eine Auflösung von 640x480 Pixeln entworfen. Da die Qualität des Fotosensors einer Videokamera die Bildverarbeitung erheblich beeinträchtigen kann, wurde während der Entwicklung eine qualitativ hochwertige Kamera vom Typ „Webcam Pro 9000“ verwendet. Das Kameramodell besitzt eine Glaslinse.

Die Anwendung wurde mithilfe der Microsoft Visual Studio 2005 Pro Entwicklungsumgebung in C++ entwickelt. Alle entwickelten Komponenten sind objektorientiert programmiert. Es wird die auf Computer Vision spezialisierte quelloffene Bibliothek OpenCV 2.0 [Cor][BK08] als Grundlage der Bildverarbeitungsprozesse verwendet. Die in dieser Arbeit verwendeten Verfahren, die in OpenCV integriert sind[6], wurden angepasst und erweitert. Durch die Bibliothek OpenMP [Boa][Cha00] wurde die Anwendung für Mehrprozessorsysteme optimiert. Die Anwendung verarbeitet das Video und speichert für jedes Bild in einer Textdatei die Position des Kopfes, die Winkel der geschätzten Kopfpose und die Größe des Kopfes in Form eines Skalierungsfaktors.

[6] Betrifft Gauss-Pyramiden und den Median-Filter (siehe Abschnitt 2.3.1), die RGB $\mapsto$ CIE-L*a*b Bildtransformation (siehe Abschnitt 2.3.1 und Anhang A), den Viola/Jones-Detektor (siehe Abschnitt 2.3.2) und den Kalman-Filter (siehe Abschnitt 2.3.3).

2.3 Kontinuierliche Kopflokalisation

Die Kopfregion wird kontinuierlich ermittelt, indem das Bild zunächst anhand von Hautfarbe in Bereiche aufgeteilt wird, in denen die Gesichtsstruktur vermutet wird (siehe Abschnitt 2.3.1). Das Gesicht wird anschließend in diesen Regionen detektiert (siehe Abschnitt 2.3.2) und unter Zuhilfenahme eines Bewegungsmodells verfolgt (siehe Abschnitt 2.3.3).

2.3.1 Hautdetektion

Menschliche Hautfarbe ist ein relativ stabiler Indikator zur effektiven Detektion von Gesichtern in Farbbildern [YKA02]. Aus diesem Grund wird das Bild zunächst in Regionen zerlegt, in denen großflächig Haut zu sehen ist. Dadurch wird der Suchbereich der angeschlossenen Gesichtsdetektion (siehe Abschnitt 2.3.2) eingegrenzt. Die Hautregionen werden durch eine *bottom-up* Verarbeitung des Eingangsbildes ermittelt. Die Reihenfolge dieser Verarbeitung wird in Abbildung 5 veranschaulicht und im Folgenden erläutert.

Vorverarbeitung Die Hautfarbenregionen werden nicht auf dem hochauflösenden RGB-Eingangsbild detektiert, sondern auf einer verkleinerten Fassung, die weiter vorverarbeitet und in den CIE-L*a*b Farbraum transformiert wird.

Gauss-Pyramide Das Eingangsbild wird zunächst in Form einer Gauss-Pyramide [AAB+84] repräsentiert. Dazu wird ausgehend vom Ursprungsbild G_0 jedes Bild G_i mit einem 5x5 Gauss-Kernel tiefpassgefiltert und jede zweite Bildzeile und -spalte entfernt, um die Bildstufe G_{i+1} der Gauss-Pyramide zu erzeugen. Pro Stufe werden zunehmend strukturelle Texturdetails entfernt. Dies steigert die Verarbeitungsgeschwindigkeit und ermöglicht die Verarbeitung grober Bildstrukturen [Bur83].

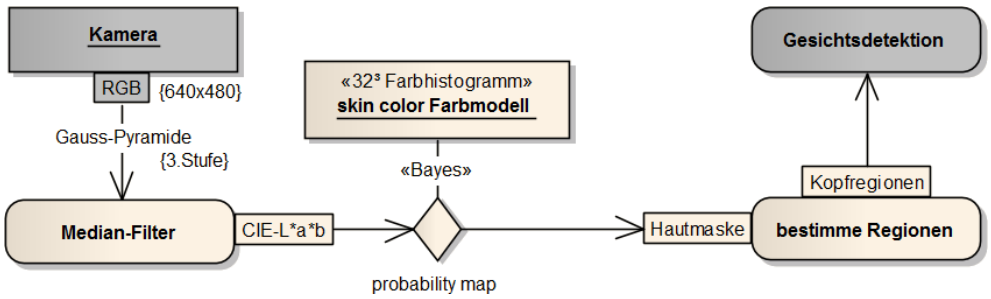


Abb. 5: Ablauf der *skin color detection*.

Median-Filter Die dritte Stufe der Gauss-Pyramide (160x120 Pixel) wird anschließend mit einem Median-Filter [Tuk77] verarbeitet. Median-Filter sind für die Unterdrückung von *salt-and-pepper noise* geeignet. Ähnlich der Tiefpassfilterung werden Texturdetails entfernt, allerdings werden die Abgrenzungen zwischen gleichmäßig gefärbten Flächen weitestgehend erhalten.

Der Median-Filter ist ein Rangordnungsfilter. Bei der Verarbeitung wird für jeden Pixel $\vec{p}$ eine Werteliste der sich im Radius r befindenden Nachbarpixel aufgestellt und aufsteigend sortiert. Anschließend wird $\vec{p}$ durch den mittleren Wert (Median) der sortierten Werteliste ersetzt. Der Algorithmus ist nicht linear, nicht separierbar und hat in seiner ursprünglichen Definition eine Laufzeit von $O(r^2 \log r)$, sodass er für Anwendungen mit zeitkritischen Anforderungen nicht geeignet ist. Hier wird ein Verfahren von Simon Perreault und Patrick Hébert [PH07] verwendet, dessen Algorithmus farbige Bilder in konstanter Laufzeit $O(1)$ unabhängig von der Wahl des Radius r verarbeiten kann.

Bestimmung der Hautmaske Die Hautregionen des Farbbildes werden durch eine binäre Hautmaske (*skin mask*) repräsentiert, die für jeden Pixel die Zugehörigkeit zur *skin-* oder *non-skin color class* ($\Omega = \{S, \neg S\}$) festlegt (vergleiche Abbildung 2(b)). Diese Zugehörigkeit wird durch eine farbbasierte Pixelklassifikation ermittelt. Dazu wird ein statistisches, nicht-parametrisches Hautfarbenmodell verwendet. Für dieses Hautfarbenmodell werden die Auftrittswahrscheinlichkeiten von *skin-* und *non-skin*-Farben unter Zuhilfenahme zweier Farbhistogramme modelliert.

Aufstellung des Farbmodells Die konditionalen Wahrscheinlichkeiten $Pr(\vec{c}|S)$ und $Pr(\vec{c}|\neg S)$ drücken die Zugehörigkeit eines Farbtupels $\vec{c}$ zur *skin-*, bzw. *non-skin color class* aus. Diese Wahrscheinlichkeiten lassen sich ermitteln, wenn für die Beobachtung eines Farbtupels dessen Klassenzugehörigkeit bekannt ist. Für alle Farbtupel der Compaq-Datenbank (siehe Abschnitt 2.1.1) wird pro Beobachtung eines Farbtupels der korrespondierende Histogrammeintrag (*bin*) $H^{(S)}(\vec{c})$, bzw. $H^{(\neg S)}(\vec{c})$ um 1 erhöht. Die Werte der *bins* werden anschließend durch die Anzahl aller Farbbeobachtungen pro Klasse geteilt [KMB07]:

$$Pr(\vec{c}|S) = \frac{H^{(S)}(\vec{c})}{|S|} \qquad Pr(\vec{c}|\neg S) = \frac{H^{(\neg S)}(\vec{c})}{|\neg S|} \tag{1}$$

Die absoluten Auftrittswahrscheinlichkeiten entsprechen dem Verhältnis der Beobachtungen einer Klasse zur Summe aller Beobachtungen [JR02]:

$$Pr(S) = \frac{|S|}{|S| + |\neg S|} \qquad Pr(\neg S) = \frac{|\neg S|}{|S| + |\neg S|} = 1 - Pr(S) \tag{2}$$

Farbbasierte Klassifikation von Pixeln Um für ein Farbtupel $\vec{c}$ eines Pixels $\vec{p}$ die Wahrscheinlichkeit $Pr(S|\vec{c})$ zu berechnen, dass es Hautfarbe repräsentiert, wird das Bayes-Theorem angewendet [VVA03]:

$$Pr(S|\vec{c}) = \frac{Pr(\vec{c}|S) \cdot Pr(S)}{Pr(\vec{c}|S) \cdot Pr(S) + Pr(\vec{c}|\neg S) \cdot Pr(\neg S)} \tag{3}$$

Wird für jeden Pixel eines Farbbildes $Pr(S|\vec{c})$ berechnet, entsteht eine Wahrscheinlichkeitskarte, die mit ihren Intensitäten die Wahrscheinlichkeit eines jeden Pixels wiedergibt, zur *skin color class* zu gehören (*skin probability map*) [QHTK02]. Eine binäre Hautmaske, die bestätigte Hautregionen repräsentiert, wird unter Zuhilfenahme eines Schwellwertes ω_S abgeschätzt, indem jeder Pixel genau dann $= 1$ gesetzt wird, wenn $Pr(S|\vec{c}) \geq \omega_S$ ist [VVA03].

CIE-L*a*b Farbraum Trotz eines breiten Spektrums möglicher Hauttöne [vL27] unterscheiden diese sich weniger in der Farbe als in der Helligkeit [GCPC95]. Für die Erkennung von Hautfarbe sind daher Farbräume geeignet, die eine getrennte Charakterisierung der Farb- und Helligkeitswerte eines Farbtupels zulassen. Der RGB-Farbraum basiert auf dem Tristimulusprinzip[7] [VH67] und ist daher nicht dazu geeignet.

Der CIE-L*a*b [Nor79] Farbraum ist ein geeigneterer Farbraum [KP92] und wird in dieser Arbeit verwendet. Er wurde zur Nachempfindung des menschlichen Farbensehens entwickelt und basiert auf der Gegenfarbentheorie[8] [Her64] (siehe Abbildung 6). Im CIE-L*a*b Farbraum sind die Abstände beliebiger Farbtupel annähernd gleichförmig, sodass die numerische Änderung einer Tupel-Komponente einer äquivalenten, empfindungsgemäßen Farbänderung entspricht. CIE-L*a*b gehört damit zur Klasse der perzeptuellen Farbräume.

Das RGB-Bild wird sowohl für das Training als auch für die *skin color detection* zunächst in den CIE-L*a*b Farbraum überführt. Die Transformation eines RGB-Farbtupels in den CIE-L*a*b Farbraum (siehe Anhang A) ist nicht

[7] RGB-Farben werden durch additive Mischung der Primärfarben Rot, Grün und Blau gebildet.

[8] Farben werden durch die separate Angabe der Helligkeit und des Verhältnisses der Gegenfarben Grün/Rot ($\pm a$) und Blau/Gelb ($\pm b$) angegeben.

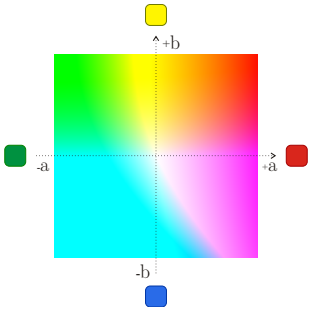


Abb. 6: Darstellung der CIE-L*a*b Chrominanzebene für $L = 100\%$. Farbe wird durch ihre relativen Gegenfarbenanteile von Grün/Rot (±a) und Blau/Gelb (±b) charakterisiert.

linear und mathematisch aufwendig (siehe Anhang A). Ein Farbhistogramm für 8-Bit CIE-L*a*b Farbtupel besteht aus 256x256x256 *bins*.

Quantisierung Farbmodelle auf Histogrammbasis sind einfach zu trainieren und schnell in der Anwendung. Allerdings haben sie einen hohen Speicherbedarf [KMB07]. Quantisiert man sie, kann man diesen reduzieren und Lücken im Histogramm schließen. Die CIE-L*a*b Farbhistogramme werden in Anlehnung an [JR02] mit dem Faktor 4 quantisiert, sodass sie 32x32x32 *bins* besitzen. Bei der Aufstellung des Modells müssen folglich die Koordinaten der *bins* ebenfalls durch Quantisierung der Farbtupelwerte bestimmt werden. Die Koordinate wird durch Multiplikation jedes Farbtupelwerts mit $s = \frac{1}{4}$ ermittelt[9]. Durch den Faktor 4 fallen auf diese Weise $4^3 = 64$ Farbtupel des originalen CIE-L*a*b Farbraums in jeweils einem *bin* zusammen. Dies entspricht einer surjektiven Abbildung [Bou04] vom Bild- in den Histogrammraum.

Segmentierung der Hautmaske Die Hautregionen werden ermittelt, indem die Zusammenhangskomponenten der *skin mask* bestimmt werden. Die Laufzeiten klassischer *region growth*-Verfahren [Z+76] sind nicht linear und für die Verarbeitung großer Bilder in Echtzeitanwendungen nur bedingt geeignet. In dieser Arbeit wird ein Verfahren von Chang et al. [CCL04] verwendet, mit dem man die Zusammenhangskomponenten eines Binärbildes durch Nachverfolgung der Konturen ermittelt. Der Algorithmus besitzt eine lineare Laufzeit. Die äußeren Grenzen der Konturen werden berechnet, um die die Hautregionen einschließenden *axis aligned bounding boxes* (AABB's) ermitteln zu können und diese an die Gesichtsdetektion (siehe Abschnitt 2.3.2) weiterzugeben.

[9] Beispiel: Der *bin* des Farbtupels $LAB = (50, 80, 70)$ hätte im mit $s = \frac{1}{4}$ quantisierten Histogramm die Koordinate $(\lfloor \frac{50}{4} \rfloor, \lfloor \frac{80}{4} \rfloor, \lfloor \frac{70}{4} \rfloor) = (12, 20, 17)$.

2.3.2 Gesichtsdetektion

Die ermittelten *axis aligned bounding boxes* (AABB's) der *skin color detection* werden dazu verwendet, das hochauflösende Bild in Bereiche aufzuteilen, in denen der Kopf potentiell abgebildet sein könnte. Das in diesen Bereichen detektierte Gesicht wird an das Tracking weitergeleitet. Für den hier verwendeten Gesichtsdetektor wird ein von Viola und Jones entwickeltes [VJ01] und durch Lienhart und Maydt erweitertes [LM02] Verfahren zur effizienten, musterbasierten Detektion von Gesichtern im Rahmen eines *sliding window approaches* [HARK96][CHLH08] eingesetzt. Der Ablauf der Gesichtsdetektion ist in Abbildung 7 schematisch dargestellt.

Gesichtsdetektion mit Haar wavelets Mit dem von Viola und Jones entwickelten *face detection*-Verfahren werden zur Detektion von Gesichtern eine Reihe von sogenannten *weak classifiern* verwendet, die mithilfe des *AdaBoost*-Verfahrens [FS97] trainiert werden. Für jeden dieser *classifier* wird das Muster eines Gesichtsmerkmals in Form eines *Haar features*[10] [CPPP98] gelernt. Lienhart und Maydt erweitern die Menge der lernbaren *Haar features* um diagonale Varianten [LM02] (siehe Abbildung 8) und ersetzen das Trainingsverfahren durch *Gentle Ada-Boost* [FS01]. Auf diese Weise konnten sie sowohl die Geschwindigkeit als auch die Detektionsgenauigkeit des Gesamtverfahrens erhöhen [LKP03][BK08].

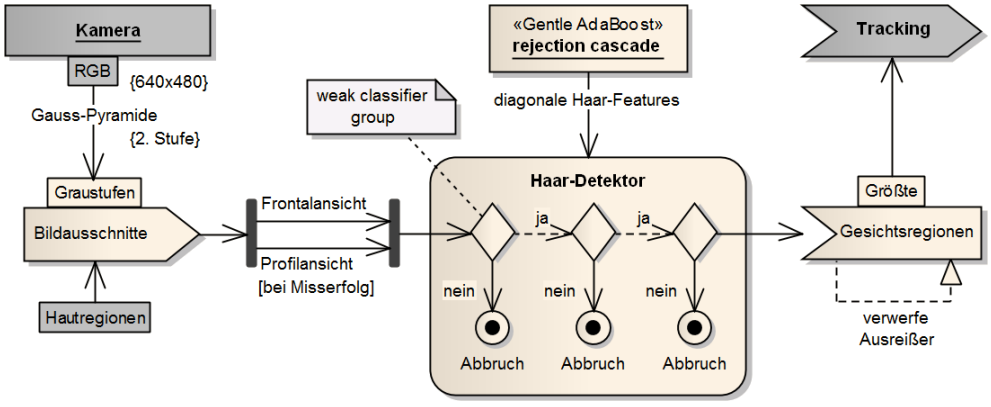


Abb. 7: Ablauf und Einordnung der *face detection*. Das *rejection cascade* Diagramm ist an [VJ01, Abbildung 4] angelehnt.

[10] Die korrekte, aber ungebräuchliche Bezeichnung dieser Muster lautet *Haar-like features*, da sie an *Haar wavelets* [Haa10][M+89] angelehnt sind. Die Bezeichnung bezieht sich auf den Entdecker des *Haar wavelets*, Alfred Haar.

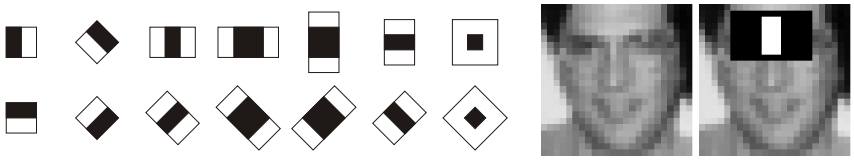


Abb. 8: *Haar features* nach [LM02] (links) und Darstellung eines *weak classifiers* der Nasenregion (rechts; modifizierte Abbildung nach [VJ01]).

Rejection cascades Die *weak classifier* sind hierarchisch in Gruppen als sogenannte *rejection cascade* organisiert. Ein Gesicht wird in einem Bildausschnitt genau dann detektiert, wenn es mit jeder *weak classifier*-Gruppe nacheinander als Gesicht klassifiziert werden kann.

Die Klassifikation einer Gruppe berechnet sich, indem die *Haar features* der *weak classifier* mit dem Bild verglichen werden. Die gewichtete Summe der *weak classifier*-Antworten dient als Entscheidungsbasis[11] der Gruppenklassifikation. Sobald eine Gruppe den Bildausschnitt als Gesicht falsifiziert, wird er frühzeitig verworfen. Die Gruppen werden ausgewählt, indem man mit der *rejection cascade* zunächst grobe Muster überprüft, die die generelle Gestalt eines Gesichts wiederspiegeln. Gegen Ende der *rejection cascade* werden *weak classifier* eingesetzt, deren Muster detailiertere Strukturen und relative Anordnungen der Gesichtsmerkmale zueinander berücksichtigen.

Verarbeitung Das Resultat der *skin color detection* sind AABB's, die angeben, wo sich im Bild großflächige Hautregionen befinden. Da *Haar features* nur für Intensitätsbilder definiert sind, wird die zweite Stufe der Gauss-Pyramide des RGB-Eingangsbildes in seine Graustufendarstellung gebracht (siehe Anhang A). Anschließend wird das Gesicht in den um 20% vergrößerten AABB's der Hautregionen in als Frontal[12]- und (bei Misserfolg) als Profilansicht gesucht. Die Detektionen werden in Form eines Quadrates angegeben.

Da dicht beieinanderliegende Detektionen ein robuster Hinweis auf ein Gesicht sind, werden solche im Rahmen dieser Arbeit zusammengefasst und Detektionen mit weniger als drei Nachbarn in unmittelbarer Nähe verworfen. Aufgrund der Anforderung, dass in der Bildsequenz nur Einzelpersonen sichtbar sind, wird die größte Detektion als erkannte Kopfregion $\vec{\Gamma} = (\Gamma_x, \Gamma_y, \Gamma_w, \Gamma_h)^T$ verwendet und an das Tracking weitergegeben.

[11] Die Gewichtungen werden, genauso wie die *Haar features* und die Gruppierung der *rejection cascade*, während des Trainings mit *AdaBoost* [FS97] gelernt.

[12] Für die Detektion von Frontalansichten wird die originale *rejection cascade* von Lienhart und Maydt verwendet. Diese wurde mit Trainingsbeispielen von einer Größe von 20x20 Pixeln trainiert und liegt der OpenCV 2.0 Distribution [Cor] bei.

2.3.3 Tracking

Unter Zuhilfenahme eines Bewegungsmodells wird die zu verfolgende Kopfregion modelliert und der kontinuierliche Bewegungsverlauf, bzw. die Größenänderung der Gesichtsregion unter Verwendung des diskreten Kalman-Filters [Kal60][BW01] geglättet.

Bewegungsmodell Die *axis aligned bounding box* $\vec{\Gamma} = (\Gamma_x, \Gamma_y, \Gamma_w, \Gamma_h)^T$ ist das Resultat der Gesichtsdetektion und entspricht der zu verfolgenden Gesichtsregion. $\vec{\Gamma}$ gibt deren Position ($\Gamma_{\{x,y\}}$), Breite (Γ_w) und Höhe (Γ_h) im Bild an. Um die Bewegung des Kopfes über die Zeit zu berücksichtigen, sind zusätzlich die linearen Änderungsraten $\Delta\Gamma_x$, $\Delta\Gamma_y$, $\Delta\Gamma_w$ und $\Delta\Gamma_h$ notwendig. Diese sind nicht messbar und versteckte Zustände des Bewegungsmodells $\vec{X}^{(\vec{\Gamma})}$:

$$\vec{X}^{(\vec{\Gamma})} = (\Gamma_x, \Gamma_y, \Gamma_w, \Gamma_h, \Delta\Gamma_x, \Delta\Gamma_y, \Delta\Gamma_w, \Delta\Gamma_h)^T \tag{4}$$

***predictor/corrector*-Struktur** Die erste Gesichtsdetektion $\vec{\Gamma}_0$ wird als initialer Zustand $\vec{X}_0^{(\vec{\Gamma})}$ verwendet. Nacheinander wird pro folgendem Zeitschritt t die Kopfregion $\vec{\Gamma}_t^*$ mithilfe von $\vec{X}_{t-1}^{(\vec{\Gamma})}$ geschätzt und an die Kopfposenerkennung weitergegeben. Anschließend wird $\vec{X}_t^{(\vec{\Gamma})}$ mit $\vec{\Gamma}_t$ korrigiert.

Dieses Prinzip nennt man *predictor/corrector*-Struktur. Die im *predictor*-Schritt vorrausgesagte *a-posteriori* [BW01] Kopfregion $\vec{\Gamma}_t^*$ wird auf Basis aller bisher beobachteten Messungen geschätzt. Der geschätzte Bewegungsverlauf ist robust gegenüber Ausreißern wie z.B. ausbleibenden oder falschen Detektionen, da diese im *corrector*-Schritt berücksichtigt werden. Diese Vorgehensweise entspricht einem Markov-Modell [Mar71] erster Ordnung.

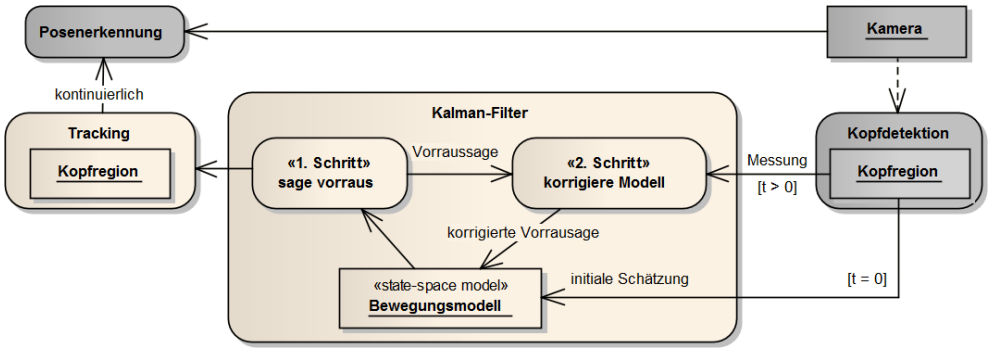


Abb. 9: *predictor/corrector*-Struktur des Kalman-Filters nach [BW01]. Das Diagramm ist von rechts nach links zu lesen.

2.4 Kopfposenerkennung

2.4.1 Posenmodelle

Die Kopfpose wird mithilfe eines *detector arrays* ermittelt [MCT09]. Es enthält mehrere Kopfposenmodelle P (*pose model*). Jedes *pose model* ist einem diskreten Kopfposenwinkel (Θ, Ψ) im Posenraum $\boldsymbol{\mathcal{P}}$ (*pose space*) zugeordnet. Ein *pose model* $P = (\boldsymbol{\mathcal{A}}, \boldsymbol{\mathcal{M}}, \boldsymbol{\mathcal{W}})$ modelliert ein Gesicht durch Landmarken $\boldsymbol{\mathcal{M}}$, Texturmodelle $\boldsymbol{\mathcal{A}}$ und Auftrittswahrscheinlichkeiten $\boldsymbol{\mathcal{W}}$ von Gesichtsmerkmalen.

$\boldsymbol{\mathcal{M}}$ bezeichnet den mittleren Modellgraphen (*mean model graph*) einer Kopfposenansicht. Jede Landmarke $\vec{l}_l \in \boldsymbol{\mathcal{M}}$ korrespondiert mit einem Gesichtsmerkmal $l \in \boldsymbol{\mathcal{F}}$ und befindet sich in einer relativen Position zum Zentrum des Graphen. Das Texturmodell $\mathcal{A}_l = (\vec{\mu_l}, \vec{\sigma_l^2}) \in \boldsymbol{\mathcal{A}}$ beschreibt das Aussehen des Gesichtsmerkmals l im Modell der Kopfposenansicht von P. Die Auftrittswahrscheinlichkeit des Gesichtsmerkmals wird durch $w_l \in \boldsymbol{\mathcal{W}}$ angegeben. Der globale *match score* [Bru09] von P an der Pixelstelle $\vec{\xi}$ (*pose score*) wird durch Aufsummierung der mit $\boldsymbol{\mathcal{W}}$ gewichteten lokalen *match scores* der Texturmodelle $\boldsymbol{\mathcal{A}}$ (*feature score*) an den Landmarken $\vec{l}'_l = \vec{\xi} + \vec{l}_l$ im Bild ermittelt (siehe Abschnitt 2.4.6).

2.4.2 Gabor wavelet transformation

Das Texturmodell jedes *pose models* P basiert auf der *Gabor wavelet transformation* (GWT) [Dau85][Wür95]. Die GWT ermöglicht die Verarbeitung visueller Reize auf mehreren Abstraktionsebenen in ähnlicher Weise, wie sie im Gehirn stattfindet [JP87]. Dies wird ermöglicht, indem mehrere durch Gaussglocken *gefensterte* komplexe *planare Wellen* in unterschiedlichen *Frequenzen* und *Richtungen* als Faltungskerne verwendet werden. Man bezeichnet diese auch als *Gabor wavelets*. *Gabor wavelets* sind komplex, mittelwertfrei und invariant gegenüber leichten Deformationen und Beleuchtungsänderungen im Bild. Die wellenförmige Ausprägung eines *Gabor wavelets* ermöglicht die strukturelle Beschreibung der Umgebung eines Pixels (siehe Abbildung 10). Nützlich für die Analyse und den Vergleich lokaler Texturmerkmale ist die Verwendung der diskreten *Gabor wavelet*-Familie [LWvdM97] bestehend aus $J = 40$ *Gabor wavelets* $\psi_{\vec{k}_j}$ in 5 Skalierungen $v = 0...4$ und 8 Rotationen $\mu = 0...7$ mit $\sigma = 2\pi$ für

$$\psi_{\vec{k}_j}(\vec{x}) = \frac{k_j^2}{\sigma^2} \cdot e^{-\frac{k_j^2 \cdot \vec{x}^2}{2\sigma^2}} \cdot \left(e^{i \vec{k}_j \vec{x}} - e^{-\frac{\sigma^2}{2}} \right) \tag{5}$$

$$\vec{k}_j = \begin{pmatrix} k_{j;x} \\ k_{j;y} \end{pmatrix} = \begin{pmatrix} k_v \cdot \cos\varphi_\mu \\ k_v \cdot \sin\varphi_\mu \end{pmatrix}, \qquad k_v = 2^{-\frac{v+2}{2}} \cdot \pi, \qquad \varphi_\mu = \mu\frac{\pi}{8} \tag{6}$$

Die Skalierung v definiert sowohl die Größe der Gaussglocke als auch die Frequenz der planaren Welle. Die Formel zur Berechnung von $\psi_{\vec{k}_j}(\vec{x})$ bezieht sich auf *Gabor wavelets* im Ortsraum. Obwohl die GWT nach einem Schema von Würtz [Wür95] auch im Frequenzraum durchführbar ist, wird hier darauf verzichtet und die Faltung eines Bildes, wie in Abbildung 10 dargestellt, im Ortsraum durchgeführt[13].

Die komplexen Antworten $\mathcal{J}_j = a_j \cdot e^{i\phi_j}$ der $J = 40$ *Gabor wavelets* $\psi_{\vec{k}_j}(\vec{x})$ $(j = 1, ..., J)$ an einer Bildstelle werden als *Gabor jet* $\vec{\mathcal{J}} = (\mathcal{J}_1, ..., \mathcal{J}_J)$ bezeichnet. Um lokale Kontraste auszugleichen, ist es hilfreich, die Absolutwerte eines *jets* zu normalisieren [GW09, (9)]:

$$a_j \leftarrow \frac{a_j}{\sqrt{\sum_{j=1}^{J} a_j^2}} \qquad j = 1, ..., J \tag{7}$$

2.4.3 Gabor-basierte Texturmodelle

Merkmalsbasierte Graphen, die ein Texturmodell auf Basis der GWT bilden, nennt man Gaborgraphen. Diese werden im Rahmen des *Elastic Bunch Graph Matchings* (EBGM) zur Gesichtserkennung verwendet [LWvdM97]. Im EBGM wird ein Texturmodell des Gesichts in Form eines *face bunch graph* $G = (\boldsymbol{\mathcal{J}}, \boldsymbol{\mathcal{L}}, \boldsymbol{\mathcal{E}})$ gebildet. Dieser wird elastisch auf das Gesicht im Bild eingepasst. Ähnlich den *appearance template methods* (siehe Abschnitt 1.2) wird als lokales Texturmodell $\boldsymbol{\mathcal{J}}_l \in \boldsymbol{\mathcal{J}}$ eines Gesichtsmerkmals l mehrere *jets* $\boldsymbol{\mathcal{J}}_l = (\vec{\mathcal{J}}_1, ..., \vec{\mathcal{J}}_n)$ als n Texturinstanzen von l gespeichert (*bunch of jets*). Neben dem Vergleich der lokalen Texturmerkmale werden beim EBGM zusätzlich die Deformationen der Kanten $\boldsymbol{\mathcal{E}}$ des eingepassten Graphen während der Berechnung des *match scores* (auch Ähnlichkeitswert, *similarity*) berücksichtigt.

Der hier verwendete Gaborgraph enthält *keine* Kanteninformationen und ist *nicht elastisch*, das heißt, er wird während des *matchings* nicht deformiert und nicht an das Gesicht angepasst. Anstatt als lokale Texturmodelle *bunches of jets* zu verwenden, wird in dieser Arbeit ein *Maximum-Likelihood* (ML) -Schätzer $\mathcal{A}_l = (\vec{\mu_l}, \vec{\sigma_l^2}) \in \boldsymbol{\mathcal{A}}$ pro Gesichtsmerkmal l eingesetzt [GW09]. Die

[13] Ein Bild wird dazu mit zwei separaten Faltungskernen gefaltet. Jeder Faltungskern entspricht dem Real- und Imaginärteil eines *Gabor wavelets*.

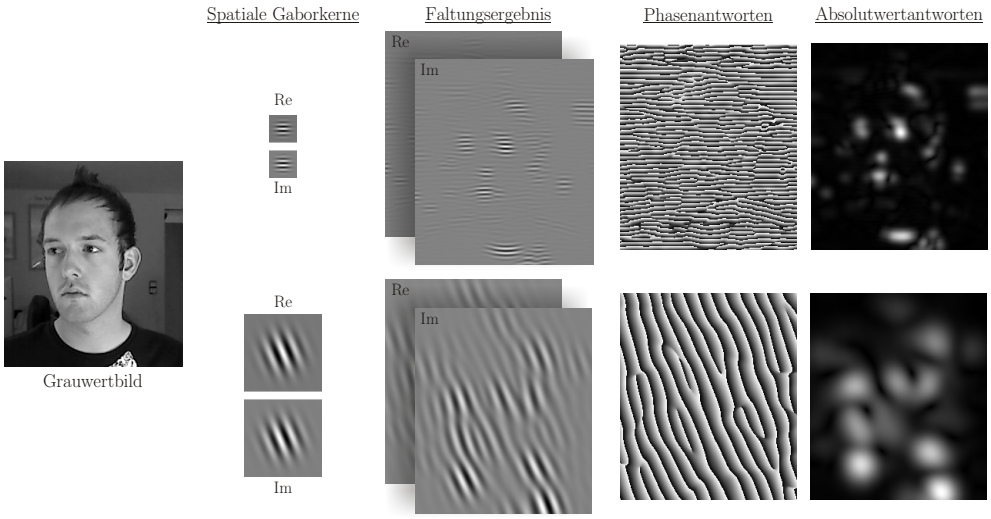


Abb. 10: Beispielhafte *Gabor wavelet transformation* eines Grauwertbildes. Abgebildet sind zwei spatiale Faltungskerne des Real- und Imaginärteils der korrespondierenden *Gabor wavelets*, die Faltungsergebnisse und die Phasen- und Absolutwertantworten. Auf Basis der Absolutwertantworten können die durch die *Gabor wavelets* repräsentierten Strukturen lokalisiert werden. Der Wertebereich aller Bilder wurde normalisiert, das Darstellungsschema ist an [LWvdM97, Abbildung 1] angelehnt.

Einträge $\vec{\mu_l} = (\mu_{l,1}, ..., \mu_{l,J})$ und $\vec{\sigma_l^2} = (\sigma_{l,1}^2, ..., \sigma_{l,J}^2)$ repräsentieren die mittleren Absolutwertantworten ($\vec{\mu_l}$) und die Varianzen ($\vec{\sigma_l^2}$) der Gabor-basierten Modelltextur des Gesichtsmerkmals l zum Vergleich mit einem *Gabor jet* eines GWT-prozessierten Bildes.

2.4.4 Training

Für das Training des *pose space* $\boldsymbol{\mathcal{P}}$ wird die komplette Pointing´04 Datenbank (siehe Abschnitt 2.1.2) verwendet. Es ergeben sich $|\boldsymbol{\mathcal{P}}| = 93$ *pose models*. Die Trainingsmenge $\boldsymbol{\mathcal{T}}$ eines *pose models* P wird in ihrem Umfang verdoppelt, indem die Gesichtssymmetrie ausgenutzt und das Bild einer jeweils horizontal entgegengesetzten Kopfpose gespiegelt wird. Es ergeben sich insgesamt $|\boldsymbol{\mathcal{T}}| = 60$ Trainingsbeispiele pro *pose model*. Es gehen auf diese Weise vier Trainingsbeispiele pro Pose und Subjekt ein.

Jedem Trainingsbild liegt ein handannotierter *pose graph* $G = (\boldsymbol{\mathcal{L}}, \boldsymbol{\mathcal{J}})$ zugrunde (siehe Abschnitt 2.1.4). An einer Landmarke $\vec{l}_l \in \boldsymbol{\mathcal{L}}$ wird eine Instanz der lokalen Textur des Gesichtsmerkmals l in Form eines *jets* $\vec{\mathcal{J}}_l \in \boldsymbol{\mathcal{J}}$ repräsentiert. Die Trainingsmenge eines *pose models* wird mit $\boldsymbol{\mathcal{T}}$ bezeichnet. Selbstokklusion kann dazu führen, dass eine Landmarke in den *pose graphs* in

$\boldsymbol{\mathcal{T}}$ nicht oder nicht durchgängig markiert ist. $\boldsymbol{\mathcal{T}}_l$ bezeichnet daher die Menge der *pose graphs* G einer Pose, in der die Landmarke $\vec{l}_l$ durchgängig markiert ist ($\boldsymbol{\mathcal{T}}_l \supseteq \boldsymbol{\mathcal{T}}$). Jedes Gesichtsmerkmal l der Landmarken $\boldsymbol{\mathcal{L}}$ stammt aus dem Gesichtsmodell $\boldsymbol{\mathcal{F}}$ (siehe Abbildung 4).

Normalisierung Um ein Texturmodell basierend auf einer GWT zu erstellen, muss sichergestellt sein, dass die zu erfassenden Texturen in einer einheitlichen Größe abgebildet sind. Andernfalls würden sich die komplexen Antworten unterschiedlich großer *Gabor wavelets* auf diesselbe strukturelle Information beziehen und indifferent in das Training des ML-Schätzers eingehen. Das statistische Modell entspräche demnach nicht mehr dem des eigentlich zu erfassenden Texturmerkmals. Deshalb müssen alle Bilder zum Trainieren eines Gabor-basierten Texturmodells so normalisiert werden, dass alle Trainingsbeispiele in einer einheitlichen Größe abgebildet sind.

Die Kopfgröße der Subjekte der Pointing´04 Datenbank variiert zwischen den Bildsequenzen, ist aber innerhalb einer Bildsequenz stabil. Pro Bildsequenz τ wird daher ein Normalisierungsfaktor $s^{(\tau)}$ berechnet, mit dem jedes Bild der Bildsequenz skaliert wird. Auf diese Weise werden die Abbildungen der Köpfe aller Kopfposenansichten auf eine einheitliche Größe gebracht. Die Berechnung von $s^{(\tau)}$ ist merkmalsbasiert. Der Abstand der Landmarken der Augenaußenwinkel ($\vec{l}_{\mathrm{CL}_{l;r}}$) im *pose graph* der Frontalansicht $G_0^{(\tau)} \leftarrow \Theta = \Psi = 0°$ wird als Maß zur Berechnung von $s^{(\tau)}$ verwendet (in Anlehnung an [GW09, Abschnitt 4.1]). In dieser Arbeit wird der Abstand der Augenaußenwinkel auf 100 Pixel als normalisierte Größe festgelegt:

$$s^{(\tau)} = \frac{100 \text{ Pixel}}{\sqrt{\left(\vec{l}_{\mathrm{CL}_l} - \vec{l}_{\mathrm{CL}_r}\right)^2} \text{ Pixel}} \tag{8}$$

Auftrittswahrscheinlichkeiten Die Wahrscheinlichkeit $w_l \in \boldsymbol{\mathcal{W}}$ des Gesichtsmerkmals l wird über das Verhältnis von $|\boldsymbol{\mathcal{T}}_l|$ zu $|\boldsymbol{\mathcal{T}}|$ berechnet:

$$w_l = \frac{|\boldsymbol{\mathcal{T}}_l|}{|\boldsymbol{\mathcal{T}}|} \tag{9}$$

Modellgraph Die Bildung des mittleren Modellgraphen (*mean model graph*) $\boldsymbol{\mathcal{M}}$ eines *pose models* P besteht aus der Bestimmung der in P sichtbaren Gesichtsmerkmale und der relativen Positionen der Landmarken.

In $\boldsymbol{\mathcal{M}}$ werden nur Landmarken derjenigen Gesichtsmerkmale l übernommen,

die häufiger als einmal in der Trainingsmenge zu sehen sind ($|\boldsymbol{\mathcal{T}}_l| > 1$). Diese Bedingung ist dadurch motiviert, dass auf diese Weise Ausreißer[14] vermieden werden und $|\boldsymbol{\mathcal{T}}_l| > 1$ eine Anforderung an das Training der ML-Schätzer ist (siehe Gleichung 11, rechts).

Um die relativen Positionen $\vec{l}_l$ der Landmarken zu bestimmen, müssen die *pose graphs* so übereinandergelegt werden, dass die Landmarken relativ nah beeinanderliegen, um ihre Ortsvektoren anschließend mitteln zu können. Da die *pose graphs* unterschiedliche Landmarken besitzen können, wird für jeden Graphen das mit $\boldsymbol{\mathcal{W}}$ gewichtete Zentrum aller Landmarken berechnet. Anschließend werden die einzelnen Posengraphen auf diesen Zentren superpositioniert. Die relative Landmarken-Position $\vec{l}_l \in \boldsymbol{\mathcal{M}}$ berechnet sich durch Mittelung der Landmarken $\vec{l}_l^{(t)}$ der Trainingsbeispiele $t \in \boldsymbol{\mathcal{T}}_l$ abzüglich der jeweiligen *pose graph*-Zentren ([GW09, (15),(16)], an variable $\boldsymbol{\mathcal{L}}$ angepasst), wie folgt:

$$\vec{l}_l = \frac{1}{|\boldsymbol{\mathcal{T}}_l|} \cdot \sum_{t=1}^{|\boldsymbol{\mathcal{T}}_l|} \left(\vec{l}_l^{(t)} - \frac{1}{\sum_{l'=1}^{|\boldsymbol{\mathcal{L}}^{(t)}|} w_{l'}} \cdot \sum_{l'=1}^{|\boldsymbol{\mathcal{L}}^{(t)}|} w_{l'} \cdot \vec{l}_{l'}^{(t)} \right) \tag{10}$$

Ermittlung des Texturmodells Das lokale Texturmodell $\mathcal{A}_l$ des Gesichtsmerkmals l eines *pose models* P wird ermittelt, indem jedes Trainingsbild $t \in \boldsymbol{\mathcal{T}}_l$ Gabor-transformiert wird und anschließend der *jet* $\vec{\mathcal{J}}_l^{(t)}$ an der Landmarke $\vec{l}_l^{(t)}$ in das Training der ML-Schätzerwerte $\mathcal{A}_l = (\vec{\mu_l}, \vec{\sigma_l^2})$ eingeht. Die Werte von $\mu_{l,j} \in \vec{\mu_l}$ und $\sigma_{l,j}^2 \in \vec{\sigma_l^2}$ werden für $j = 1, ..., J$ wie folgt berechnet (nach [GW09, (13),(14)], an variable Trainingsmengen angepasst):

$$\mu_{l,j} = \frac{1}{|\boldsymbol{\mathcal{T}}_l|} \cdot \sum_{t=1}^{|\boldsymbol{\mathcal{T}}_l|} a_{l,j}^{(t)} \qquad \sigma_{l,j}^2 = \frac{1}{|\boldsymbol{\mathcal{T}}_l| - 1} \cdot \sum_{t=1}^{|\boldsymbol{\mathcal{T}}_l|} \left(a_{l,j}^{(t)} - \mu_{l,j} \right)^2 \tag{11}$$

2.4.5 Normalisierung der Gesichtsregion

Die Bewegung des Kopfes in der Tiefe wird bei Verwendung des Verfahrens von McKenna und Gong nicht erkannt (siehe Abschnitt 1.3). Sie schlagen vor, das Bild in mehreren Skalierungen zu verarbeiten. Stattdessen wird hier der Normalisierungsfaktor s_γ des Eingangsbildes direkt über die Ausmaße der verfolgten Gesichtsregion $\vec{\Gamma}^*$ berechnet (siehe Abschnitt 2.3.3). Da der Normalisierungsfaktor $s^{(\tau)}$ im Training *merkmalsbasiert* über den Abstand der Augenaußenwinkel (kurz: BB, *biectocanthus breadth*) berechnet wird, muss ein

[14] Dies betrifft z.B. während der Annotation fälschlicherweise markierte Landmarken oder irreguläre Kopfhaltungen eines Subjektes im Trainingsbild.

Skalierungsfaktor s_γ *indirekt* über die Kopfbreite (kurz: HB, *head breadth*) Γ_w^* berechnet werden.

s_γ drückt aus, um wieviel Prozent das Bild vergrößert werden muss, damit der Abstand der Augenaußenwinkel = 100 Pixel wird. Dieser Faktor kann berechnet werden, wenn bekannt ist, wie weit die Augenaußenwinkel im Bild voneinander entfernt sind. Da nur Γ_w^* bekannt ist, kann diese Entfernung abgeschätzt werden, wenn bekannt ist, in welchem Verhältnis der statistische geschlechterübergreifende mittlere Augenaußenwinkelabstand $\widehat{\mu}^{(\mathrm{BB})}$ zur statistischen geschlechterübergreifenden mittleren Kopfbreite $\widehat{\mu}^{(\mathrm{HB})}$ steht:

$$s_\gamma = \frac{100 \text{ Pixel}}{\Gamma_w^* \cdot \frac{\widehat{\mu}^{(\mathrm{BB})}}{\widehat{\mu}^{(\mathrm{HB})}}} \tag{12}$$

Zur Bestimmung von $\widehat{\mu}^{(\mathrm{HB})}$ und $\widehat{\mu}^{(\mathrm{BB})}$ werden anthropometrische Messdaten verwendet [You93, Tabelle 3 und 6]. Diese sind nach Geschlechtern $(m; f)$ getrennt und hier mit $\mu_{m;f}^{\mathrm{HB}}$, bzw. $\mu_{m;f}^{\mathrm{BB}}$ bezeichnet. Sie werden in Tabelle 2 zusammengefasst und im Detail in Tabelle 5 und Tabelle 6 aufgelistet. Die Berechnung eines geschlechterübergreifenden Mittelwertes ergibt sich durch Mittelung der geschlechterspezifischen Mittelwerte und der Gewichtung anhand der Anzahl der an den Messungen beteiligten Versuchsteilnehmern:

$$\widehat{\mu} = \frac{|\mu_m| \cdot \mu_m + |\mu_f| \cdot \mu_f}{|\mu_m| + |\mu_f|} \tag{13}$$

Es ergibt sich für $\widehat{\mu}^{(\mathrm{HB})}$ und $\widehat{\mu}^{(\mathrm{BB})}$ (vergleiche Tabelle 2):

$$\widehat{\mu}^{(\mathrm{HB})} = \frac{167 \cdot 6,00 + 169 \cdot 5,74}{167 + 169} \approx 5,869 \text{ inch} \tag{14}$$

$$\widehat{\mu}^{(\mathrm{BB})} = \frac{101 \cdot 3,60 + 134 \cdot 3,49}{101 + 134} \approx 3,537 \text{ inch} \tag{15}$$

	Kopfbreite		**Augenaußenwinkelabstand**	
	männlich	*weiblich*	*männlich*	*weiblich*
Anzahl Subjekte	167	169	101	134
Mittelwert in *inch*	6,00	5,74	3,60	3,49

Tab. 2: Mittlere statistische Kopfbreite und Augenaußenwinkelabstände (nach Geschlechtern getrennt) ([You93], Zusammenfassung; die vollständigen Datenblätter befinden sich im Anhang B).

Das Verhältnis des mittleren Abstandes der Augenaußenwinkel zur mittleren Kopfbreite ist $\widehat{\mu}^{(\mathrm{BB})}/\widehat{\mu}^{(\mathrm{HB})} \approx 3,537/5,869 \approx 60,27\%$. Es ergibt sich für s_γ:

$$s_\gamma = \frac{100 \text{ Pixel}}{\Gamma_w^* \cdot \frac{\widehat{\mu}^{(\mathrm{BB})}}{\widehat{\mu}^{(\mathrm{HB})}}} \approx \frac{165,925 \text{ Pixel}}{\Gamma_w^*} \tag{16}$$

Nach der Normalisierung des Eingangsbildes wird um die Mitte des Kopfes ein 400x400 Pixel großer Bildausschnitt extrahiert und in Graustufen umgewandelt. Die Erkennung der Kopfpose findet auf diesem Bild statt.

2.4.6 Bestimmung der Kopfpose

Die Kopfpose wird ermittelt, indem mithilfe eines *pose space models* $\boldsymbol{\mathcal{P}}$ die jeweilige Kopfposenansicht im Bild detektiert wird. Dazu wird für jede Bildposition der *match score* jedes *pose models* $P \in \boldsymbol{\mathcal{P}}$ berechnet (*pose score*). Das Ergebnis der Kopfposenerkennung besteht in den Winkeln derjenigen diskreten Kopfpose, mit deren *pose model* der größte *pose score* erzielt werden kann.

Ähnlichkeitsfunktion Der *pose score* eines *pose models* P an einer Bildstelle $\vec{\xi}$ wird mithilfe der Ähnlichkeitsfunktion $S^{(P)}(\boldsymbol{\mathcal{J}})$ berechnet. Dazu wird die Abweichung der Bild-*jets* $\boldsymbol{\mathcal{J}}$, auf denen die Landmarken von $\boldsymbol{\mathcal{M}}$ bezogen auf $\vec{\xi}$ liegen, zum Texturmodell $\boldsymbol{\mathcal{A}}$ berechnet. Die Ähnlichkeit $S^{(\vec{\mu_l},\vec{\sigma_l^2})}(\vec{\mathcal{J}})$ eines Bild-*jets* $\vec{\mathcal{J}}$ zum Texturmodell $\mathcal{A}_l = (\vec{\mu_l}, \vec{\sigma_l^2})$ der Landmarke $\vec{l}_l$ wird wie folgt ermittelt (nach [GW09, (19)], gekürzt):

$$S^{(\vec{\mu_l},\vec{\sigma_l^2})}(\vec{\mathcal{J}}) = -\sum_{j=1}^{J} \frac{(a_j - \mu_j)^2}{\sigma_j^2} \tag{17}$$

Durch den negativen Faktor ist die Ähnlichkeitsfunktion als Energiefunktion zu verstehen. Die Ähnlichkeit des vollständigen *pose models* zum Bild definiert sich durch die mit $\boldsymbol{\mathcal{W}}$ gewichteten *feature scores* (nach [GW09, (19)], erweitert um Gewichte):

$$S^{(P)}(\boldsymbol{\mathcal{J}}) = -\frac{1}{\sum_{l=1}^{|\boldsymbol{\mathcal{M}}|} w_l} \cdot \sum_{l=1}^{|\boldsymbol{\mathcal{M}}|} \left(w_l \cdot \sum_{j=1}^{J} \frac{(a_{l,j} - \mu_{l,j})^2}{\sigma_{l,j}^2} \right) \tag{18}$$

Die Division durch die Summe der Gewichte normalisiert den Ähnlichkeitswert. Da dies für alle *pose models* auf diese Weise geschieht, sind die Ähn-

lichkeitswerte unabhängig von der Landmarkenanzahl und den Auftrittswahrscheinlichkeiten vergleichbar. Da der Ähnlichkeitswert beliebig gering ausfallen kann, entspricht er der unnormalisierten Wahrscheinlichkeit (*likelihood*), dass das Texturmodell des *pose models* mit der Bildstelle übereinstimmt.

Posendetektion Mit jedem *pose model* P werden an jeder Stelle $\vec{\xi}$ des Bildes die *jets* $\boldsymbol{\mathcal{J}}^{(P)}(\vec{\xi})$ ermittelt, auf denen die Landmarken des Graphen $\boldsymbol{\mathcal{M}}$ bezogen auf $\vec{\xi}$ liegen. $\vec{\xi}^{*(P)}$ ist diejenige Stelle im Bild, an der der *pose score* von P maximal ist (nach [GW09, (17)], angepasst):

$$\vec{\xi}^{*(P)} = \arg\max_{\vec{\xi}} \left\{ S^{(P)} \left(\boldsymbol{\mathcal{J}}^{(P)}(\vec{\xi}) \right) \right\} \tag{19}$$

Anstatt in einem Durchlauf das *pose model* gegen jede Bildstelle zu prüfen (*exhaustive search*), wird das Bild, wie in [LWvdM97] vorgeschlagen, erst in einem groben Raster abgesucht. Hier wird ein Abstand von $\epsilon = 8$ Pixel verwendet. Die Stelle $\vec{\xi}^{\epsilon}$ bezeichnet diejenige Rasterkoordinate im Bild, an der der *pose score* maximal ist. In einem um $\vec{\xi}^{\epsilon}$ zentrierten, 2ϵ breiten quadratischen Bereich, wird anschließend an jeder Pixelstelle das Bild erneut mit dem *pose model* verglichen, um $\vec{\xi}^{*(P)}$ zu ermitteln (siehe Abbildung 20).

Posenerkennung Es sind diejenigen Winkel das Ergebnis der Kopfposenerkennung, deren zugeordnetes *pose model* P^* den größten *pose score* im gesamten *pose space* $\boldsymbol{\mathcal{P}}$ erzielen kann:

$$(\Theta, \Psi) \leftarrow P^* = \arg\max_{P \in \boldsymbol{\mathcal{P}}} \left\{ S^{(P)} \left(\boldsymbol{\mathcal{J}}^{(P)}(\vec{\xi}^{*(P)}) \right) \right\} \tag{20}$$

Werden die jeweils höchsten *pose scores* eines jeden *pose models* in einer Matrix eingetragen und normalisiert, ergibt sich eine Wahrscheinlichkeitskarte der jeweiligen Kopfposen (*pose probability map*, vergleiche Abbildungen 21 und 22).

Glättung der Posenwinkel Durch Verwendung eines *detector arrays* können nur diskrete Kopfposenwinkel registriert werden. In Anlehnung an Abschnitt 2.3.3 wird unter Zuhilfenahme eines Kalman-Filters mit dem Bewegungsmodell $\vec{X}^{(\boldsymbol{\mathcal{P}})} = (\Theta, \Psi, \Delta\Theta, \Delta\Psi)$ die Rotation des Kopfes modelliert und die erkannten Kopfposenwinkel geglättet.

2.5 Versuchsaufbau

Im Mittelpunkt der Evaluation steht das Verfahren zur Kopfposenerkennung. Die Evaluation der Kopflokalisation beschränkt sich auf die Dokumentation der Ergebnisse der miteinander kombinierten Verfahren.

Hautfarbendetektion Die *skin probability maps* von Bildern mit natürlicher und künstlicher Beleuchtung werden unter Verwendung von quantisierten und nicht-quantisierten CIE-L*a*b Farbhistogrammen verglichen. Es werden mehrere Hautmasken durch Variierung des Schwellwerts ω_S erzeugt und die klassifizierten Hautregionen untersucht.

Gesichtsdetektion Es wird untersucht, welche AABB's der *skin color detection* an die Gesichtsdetektion weitergegeben werden und ob der verwendete Viola/Jones-Detektor in diesen Regionen Frontal- und Profilansichten des Gesichts detektieren kann. Der Viola/Jones-Detektor wird mit mehreren Kopfposenansichten geprüft, um zu untersuchen, in welchen Situationen fehlerhafte oder ausbleibende Detektionen auftreten können.

Texturmodelle Die Gabor-basierten Texturmodelle der *pose models* werden nach dem Verfahren von Günther und Würtz [GW10] einmal mit und einmal ohne Berücksichtigung der Auftrittswahrscheinlichkeiten rekonstruiert. Dies entspricht der Visualisierung derjenigen Textur, für die das Texturmodell den größten Ähnlichkeitswert erzielen würde. Es wird untersucht, inwiefern die Texturmodelle personenunabhängig sind und ob durch Verwendung von Auftrittswahrscheinlichkeiten instabile Gesichtsmerkmale weniger berücksichtigt werden.

Posenerkennung Um mit anderen Verfahren vergleichbar zu sein, werden in dieser Arbeit die in [MCT09] verglichenen, mittleren absoluten Winkelfehler und die mittleren Klassifikationsraten der Kopfposenerkennung getrennt nach Nick- und Gierwinkel evaluiert. Für die Evaluation des kompletten Verfahrens inklusive der regionenbasierten Normalisierung und separaten Kopflokalisation liegen keine Bildsequenzen mit *ground truth*-Daten vor, sodass die Kopfposenerkennung lediglich bildbasiert auf der Pointing'04 Datenbank evaluiert werden kann. Die Messergebnisse basieren ähnlich dem Training daher auf der Normalisierung anhand der Augenaußenwinkel der Frontalansicht einer Bildsequenz (vergleiche Abschnitt 2.4.4).

Zunächst wird der Trainingsfehler einmal mit und einmal ohne Berücksichtigung der Auftrittswahrscheinlichkeiten berechnet (*baseline test*). Anhand einer Zufalls-Kreuzvalidierung von 100 Testläufen (*repeated random sub-sampling cross validation*) wird auf 80% der Trainings- und 20% der Testmenge (getrennt nach Subjekten) unter Berücksichtigung von Auftrittswahrscheinlichkeiten die Leistungsfähigkeit des Verfahrens ermittelt.

Während der Kreuzvalidierung werden zusätzlich die mittleren *pose probability maps* gemittelt. Um die Klassifikationsraten interpretieren zu können, werden Histogramme für die Gier- und Nickwinkelachse erstellt, in deren *bins* erfolgreiche Klassifikationen aufaddiert und anschließend gemittelt werden.

3 Ergebnisse

3.1 Hautfarbendetektion

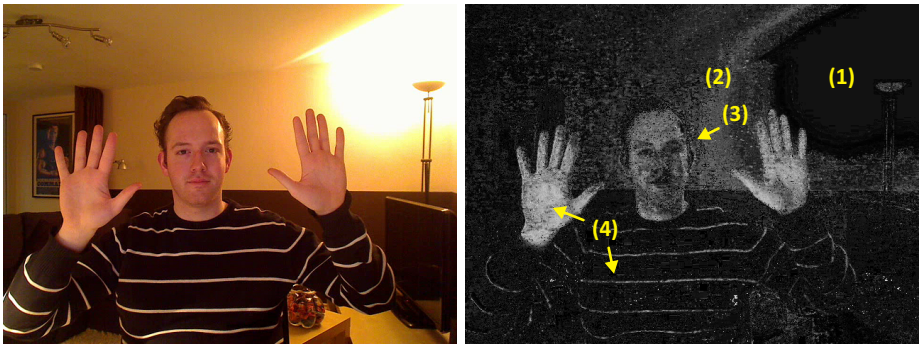
Variable Lichtverhältnisse Bei Verwendung eines nicht quantisierten Farbmodells sind die Hautflächen in der *skin probability map* des natürlich beleuchteten Bildes in Abbildung 11(a) durch eine scharf begrenzte Fläche und hohe Wahrscheinlichkeitswerten vom Hintergrund separiert. Helle Hautstellen sind in der *skin probability map* durchlöchert (siehe Abbildung 11(b)). Helle Streifen in der Kleidung werden als Haut maskiert. Der Schattenwurf des Gesichts ist in der *skin probability map* sichtbar.

Bei künstlichem, gelblichen Licht wie in Abbildung 11(c), sind die Hautflächen in der *skin probability map* verrauscht (siehe Abbildung 11(d)) und Hintergrundstrukturen sind erkennbar. Der Übergang des Scheins einer Lichtquelle im Hintergrund erzeugt eine ringförmige Region in der *skin probability map* mit hohen Wahrscheinlichkeitswerten. Die Streifen in der Kleidung haben im Vergleich zu Abbildung 11(b) erhöhte Wahrscheinlichkeitswerte.

Quantisierung Unter Verwendung von mit $s = \frac{1}{4}$ quantisierten Farbhistogrammen ist zu beobachten, dass die Verteilung der Wahrscheinlichkeitswerte in den zusammenhängenden Hautregionen der *skin probability maps* in Abbildung 12 gleichmäßiger ist. Allerdings ist auch zu beobachten, dass Teile dieser Regionen mit niedrigen Wahrscheinlichkeitswerten besetzt werden und zerfallen können.

Klassifikation Bei Verwendung unterschiedlicher Schwellwerte zur Klassifikation der *skin probability maps* von Abbildung 11(a) und Abbildung 11(c) ist zu beobachten, dass die Hautflächen im Bild mit natürlicher Beleuchtung im Vergleich zum Bild mit künstlichem, gelblichen Licht bereits bei bei niedrigen Schwellwerten klassifiziert werden können (vergleiche Abbildung 13). Auch bei Erhöhung des Schwellwerts werden für Abbildung 11(a) nur sehr wenige Hintergrundpixel klassifiziert. Für Abbildung 11(c) tritt die Klassifikation von Hautflächen erst bei vergleichbar höheren Schwellwerten ein und der Anteil des klassifizierten Hintergrundes ist wesentlich höher, als bei einem natürlich beleuchteten Bild.

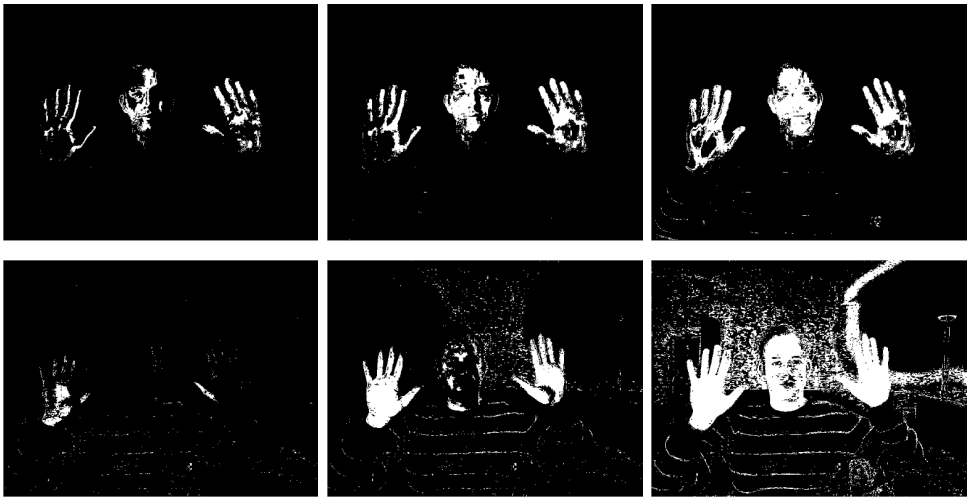
(a) natürliche Beleuchtung (b) *skin probability map*



(c) künstliche, gelbliche Beleuchtung (d) *skin probability map*

Abb. 11: *Skin probability maps* von Bildern mit unterschiedlichen Lichtverhältnissen auf Basis eines unquantisierten Farbmodells. Anmerkungen zu Abbildung 11(b): 1.) durchlöcherte, helle Hautflächen, 2.) sichtbarer Schattenwurf, 3.) helle Grauwerte werden als Haut erkannt. Anmerkungen zu Abbildung 11(d): 1.) heller Lichtschein wird nicht klassifiziert, 2.) verrauschter Hintergrund, Strukturen sichtbar, 3.) Gesicht stark verrauscht, 4) Hände mit höherer Wahrscheinlichkeitsverteilung registriert als das Gesicht, Teile der Kleidung als Hautfarbe erkannt.

Abb. 12: *Skin probability maps*, die unter Verwendung eines quantisierten und nichtquantisierten Farbmodells ($s = \frac{1}{4}$) erzeugt wurden. Jeweils links das Farbbild, daneben die *skin probability maps* unter Verwendung eines unquantisierten Farbmodells (jeweils in der Mitte) und einem quantisierten Farbmodell (jeweils rechts).



(a) $\omega_S = 20\%$ (b) $\omega_S = 40\%$ (c) $\omega_S = 60\%$

Abb. 13: *Skin masks* unter Variierung des Klassifikations-Schwellwertes und unterschiedlichen Lichtbedingungen. In der oberen Bildreihe wurde die *skin probability map* der Abbildung 11(a) und in der unteren Bildreihe die der Abbildung 11(c) mit den Schwellwerten $\omega_S = \{20\%, 40\%, 60\%\}$ (v.l.n.r) klassifiziert.

3.2 Gesichtsdetektion

Durch Hautregionen eingegrenzte Gesichtsdetektion Die *skin probability maps* der Abbildungen 11(a) und 11(c) wurden mit einem Schwellwert von $\omega_S = 20\%$ klassifiziert und in Hautregionen zerlegt. Die um 20% vergrößerten *axis aligned bounding boxes* (AABB's) dieser Hautregionen sind in Abbildung 14 dargestellt. In diesen Bereichen hat der Viola/Jones-Detektor erfolgreich die Frontalansicht detektieren können, auch wenn die AABB der Gesichtsdetektion in Abbildung 14(b) einen Teil des Hintergrunds abdeckt. Es ist zu beobachten, dass bei natürlicher Beleuchtung die Hautregionen erfolgreich registriert werden, während dies bei künstlicher, gelblicher Beleuchtung nicht der Fall ist (vergleiche Abbildung 14(a) mit Abbildung 14(b)).

Detektion rotierter Gesichter Sowohl in Frontal-, Vollprofil-, und Halbprofilansichten decken die AABB's der *face detection* das Gesicht (vergleiche Abbildung 15, oben) ab. Bei starken Rotationen in der Bildebene oder Ansichten, in denen das Gesicht nicht mehr vollständig zu sehen ist (Blick entlang der Körperachse, $\Theta = \pm 90°$), wird kein Gesicht detektiert. In Profilansichten sind Detektionen zu beobachten, die das Gesicht nicht genau abdecken oder nur einen kleinen Teil davon.

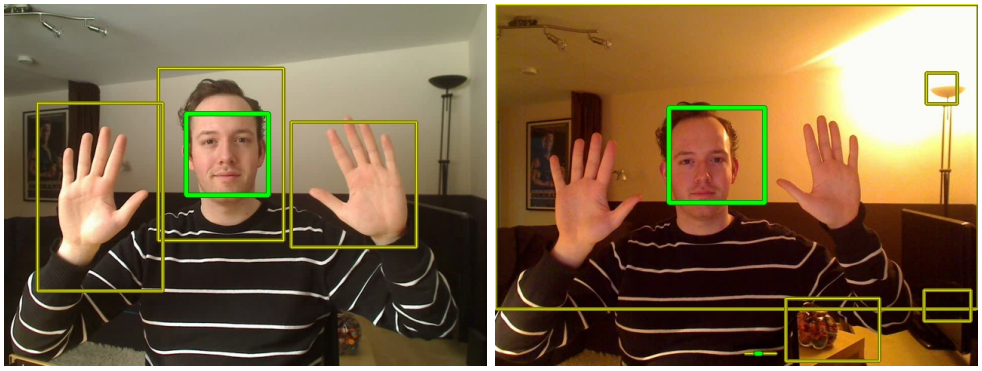


Abb. 14: Um 20% vergrößerte *axis aligned bounding boxes* (AABB's) der *skin color detection* (gelb) und AABB als Ergebnis der *face detection* mit dem Viola/Jones-Detektor (grün).

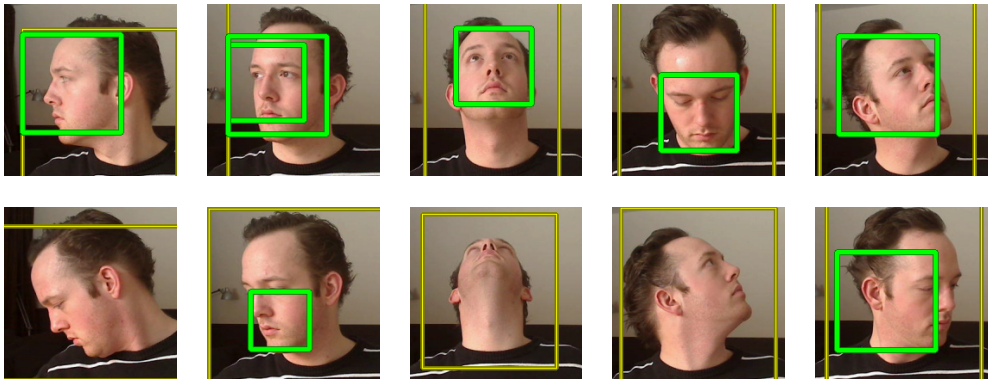


Abb. 15: Mehrere Kopfposenansichten und Antworten des Viola/Jones-Detektors (grün). Obere Reihe: korrekte Detektionen. Untere Reihe: ausbleibende oder fehlerhafte Detektionen.

3.3 Kopfposenerkennung

3.3.1 Rekonstruierte Texturmodelle

Die Rekonstruktionen [GW10] der auf der gesamten Pointing´04 Datenbank trainierten Texturmodelle ähneln menschlichen Gesichtern (siehe Abbildung 16). Die Gesichtszüge sind überwiegend männlich und weisen einen neutralen Gesichtsausdruck auf. Die Rekonstruktionen weisen keine fremdartigen Objekte wie Brillen, Schmuck, Kleidung, etc. auf und modellieren bis auf die Augenbrauen keine Gesichtsbehaarung. Die Pupille ist durchgängig erkennbar. Die Verwerfung nicht stabiler Gesichtsmerkmale aufgrund niedriger Auftrittswahrscheinlichkeiten ist besonders in Halbprofilansichten zu beobachten.

3.3.2 Baseline-Test und Kreuzvalidierung

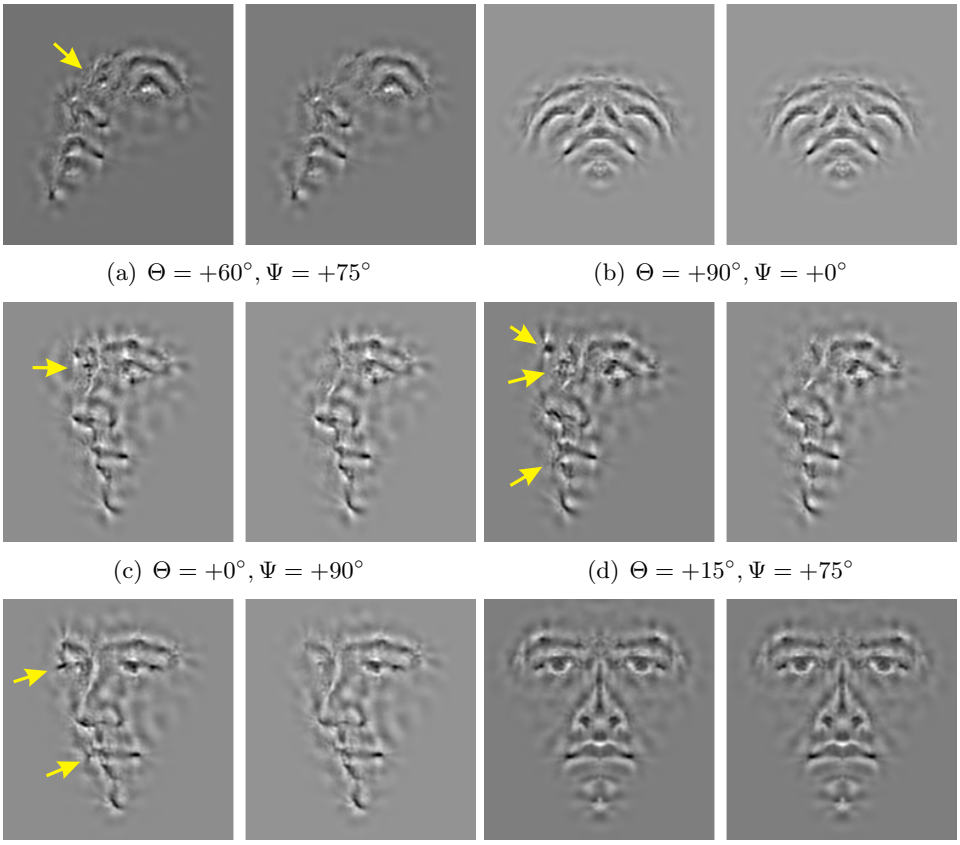
Für den *baseline*-Test wurden die *pose models* auf der kompletten Trainingsmenge trainiert (5580 Trainingsbeispiele) und auf derselben getestet, einmal mit und einmal ohne Gewichtungen. Für die Kreuzvalidierung wurde die Pointing´04 Datenbank 100 Mal auf zufällige Weise in 80% Trainings- und 20% Testmenge aufgeteilt, pro Durchlauf auf dem Trainingsset trainiert und auf dem Testset evaluiert (*repeated random sub-sampling cross validation*). Die Aufteilung basiert auf der Trennung nach Subjekten.

Es wurden drei Mittelklasse-PC's verwendet[15]. Der *baseline*-Test benötigte 18 Stunden für den Test mit Gewichten und 11 Stunden für den Test ohne Gewichte. Die Kreuzvalidierung des Verfahrens hat 8 Tage gedauert, davon wurden 2 Tage für das Training verwendet.

Baseline-Test Der mittlere absolute Nickwinkelfehler beträgt $7,82°$ ($12,99°$ ohne Gewichte) und der mittlere absolute Gierwinkelfehler $8,25°$ ($14,65°$). Die mittlere Klassifikationsrate des Nickwinkels beträgt $71,70\%$ ($57,72\%$) und die des Gierwinkels $59,55\%$ ($45,77\%$). Die Werte sind in Tabelle 3 mit den Messergebnissen der Kreuzvalidierung (siehe unten) zusammengefasst.

Kreuzvalidierung Der mittlere absolute Nickwinkelfehler liegt bei $8,826°$ (Standardabweichung $= 1,665°$, mittlerer Fehler $= 0,167°$) und der mittlere absolute Gierwinkelfehler bei $8,826°$ (Standardabweichung $= 2,180°$, mittlerer Fehler $= 0,218°$). Der Nickwinkel wurde im Mittel in $66,23\%$ aller Fälle

[15]Intel Pentium Dual 2 GHz, 1 GB RAM; Intel Dual Core2 Quad 2,34 GHz, 2,75 GB RAM, Intel Core2 Duo 1,4 GHz 4 GB RAM.



(a) $\Theta = +60°, \Psi = +75°$ (b) $\Theta = +90°, \Psi = +0°$

(c) $\Theta = +0°, \Psi = +90°$ (d) $\Theta = +15°, \Psi = +75°$

(e) $\Theta = +15°, \Psi = +60°$ (f) $\Theta = +0°, \Psi = +0°$

Abb. 16: Rekonstruierte Texturmodelle ausgewählter *pose models* nach einem Verfahren von Günther und Würtz [GW10]. Jede Unterabbildung stellt eine Rekonstruktion mit (jeweils links) und ohne Gewichtung der Gesichtsmerkmale dar (jeweils rechts). Besonders in Halbprofilen sind instabile Gesichtsmerkmale entlang der Gesichtskontur erkennbar (jeweils durch gelbe Pfeile markiert).

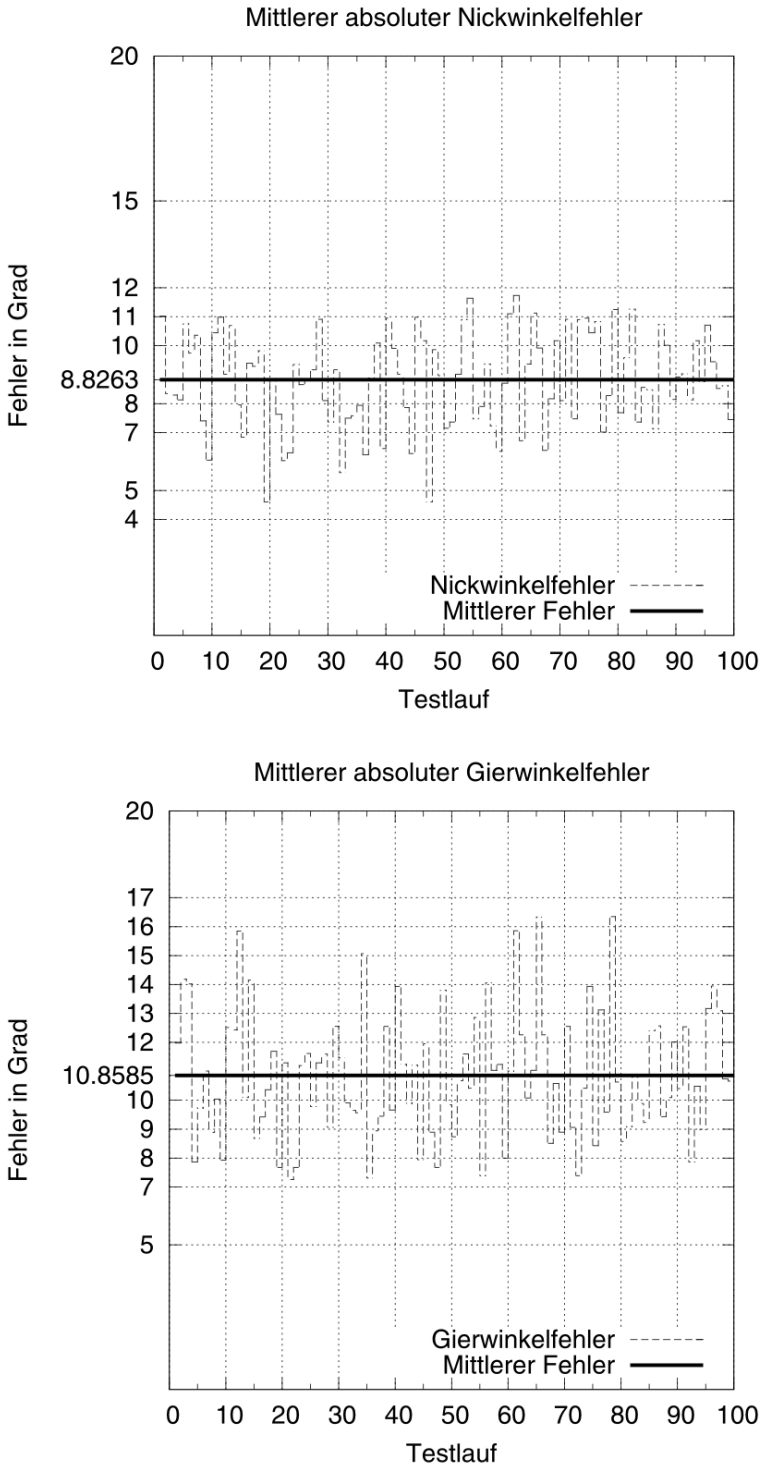


Abb. 17: Mittlere absolute Nick- (oben) und Gierwinkelfehler (unten) der Kreuzvalidierung. Die gestrichelte Linie entspricht den Werten der Testläufe, die durchgezogene Linie entspricht dem Mittelwert.

erfolgreich klassifiziert und der Gierwinkel in $50,96\%$ aller Fälle. Die Werte sind zusammen mit den Messergebnissen des Baseline-Tests (siehe oben) in Tabelle 3 zusammengefasst, die Einzelergebnisse der Kreuzvalidierung geben die Abbildungen 17 und 18 wieder. Die mittleren Klassifikationsraten der Nickwinkel $\Theta = \{-90°, 0°, +15°\}$ (siehe Abbildung 19) liegen unter dem Mittelwert von $62,23\%$. Die höchsten mittleren Klassifikationsraten sind für $\Theta = \pm 60°$ zu beobachten. Die mittleren Klassifikationsraten des Gierwinkels liegen für $\Psi = \{0°, \pm 15°, \pm 90°\}$ am höchsten, die niedrigsten Klassifikationsraten sind für $\Psi = \pm 75°$ zu beobachten (vergleiche Abbildung 19(b)).

Ursachen fehlerhafter Kopfposenschätzungen Es kann beobachtet werden, dass Texturmodelle von Profilansichten sichelförmig sind. Bei einem Vergleich mit einer Frontalansicht werden diese sichelförmigen Modelle auf einer Gesichtshälfte gefunden und erzielen einen ähnlich hohen *pose score* wie ein Texturmodell einer Frontalansicht (siehe Abbildung 20). Künstlich verdeckte Gesichtsmerkmale[16] (siehe Abbildung 21(c)) können dazu führen, das ein *pose model*, das diese weniger berücksichtigt, einen höheren *pose score* besitzt, als ein vergleichbares Modell, dass dieselben Gesichtsmerkmale mit höheren Gewichten berücksichtigt. Silhouetten, die zufällig durch Lichteinfall, Kleidung usw. entstehen, beeinflussen das Matching auf dieselbe Weise (siehe Abbildungen 21(e) und 21(f)). Da Graphen nur an den Stellen geprüft werden, an denen alle Landmarken innerhalb des Bildes liegen (vergleiche Abbildung 20), kann die Schätzung für Gesichter in der Nähe oder außerhalb des Bildrandes (vergleiche Abbildungen 21(d) und 21(e)) ebenfalls fehlerhaft sein.

Pose probability maps Das mittlere globale Minimum aller *pose scores* ist $S_{\min} = -168,701$ und das Maximum $S_{\max} = -49,7592$. Die in der Kreuzvalidierung gemittelten *pose probability maps* wurden anhand dieser Werte normalisiert.

Im Gegensatz zu Profilansichten ist besonders für Frontalansichten zu beobachten, dass viele Kopfposenmodelle ähnlich hohe *pose scores* erzielen. Ist der Kopf zur Seite gedreht, sind die *scores* von Posen entgegengesetzter Gierwinkel trotz der unterschiedlichen Kopfrichtung relativ hoch. Das Gefälle der *pose scores* ist für Vollprofilansichten und abwärts gerichtete Nickwinkel in Relation zu Frontalansichten am größten.

[16] Beispielsweise durch Gesichtsbehaarung, Brillen, Schmuck oder Kleidung.

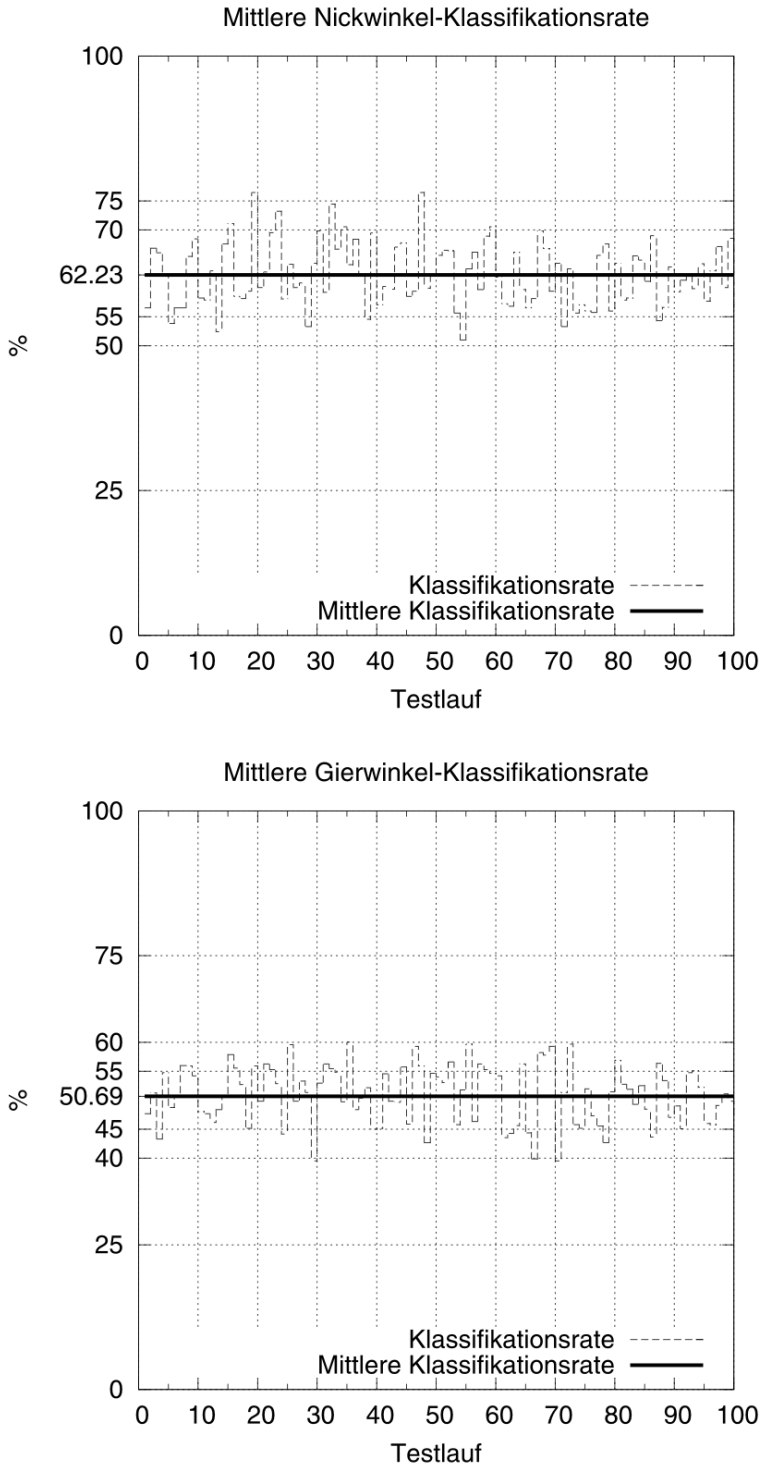


Abb. 18: Mittlere Nick- (oben) und Gierwinkelklassifikationsraten (unten) der Kreuzvalidierung. Die gestrichelte Linie entspricht den Werten der Testläufe, die durchgezogene Linie entspricht dem Mittelwert.

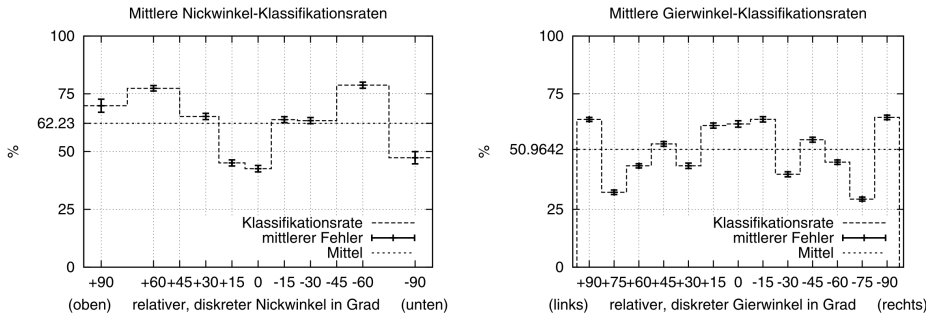


Abb. 19: Mittlere Klassifikationsraten, aufgeschlüsselt nach diskreten Posenwinkeln.

	Mittl. abs. Winkelfehler		**Mittl. Klassifikationsrate**	
Evaluationsverfahren	Nickwinkel	Gierwinkel	Nickwinkel	Gierwinkel
Baselinetest				
$\mu^{(\boldsymbol{\mathcal{W}})}$	7,82°	8,25°	71,70%	45,77%
$\mu^{(\boldsymbol{\mathcal{W}}')}$	12,99°	14,56°	57,72%	59,55%
Kreuzvalidierung				
$\mu^{(\boldsymbol{\mathcal{W}})}$	8,83°	10,86°	62,23%	50,96%
$\sigma^{(\boldsymbol{\mathcal{W}})}$	1,67°	2,18°	5,52%	5,05%
$\sigma_{\mu}^{(\boldsymbol{\mathcal{W}})}$	0,17°	0,22°	0,55%	0,50%

Tab. 3: Mittlere absolute Winkelfehler und Klassifikationsraten des Baseline- und Kreuzvalidierungstestes. Die Angabe von $\boldsymbol{\mathcal{W}}$ oder $\boldsymbol{\mathcal{W}}'$ bezieht sich auf die Berücksichtigung von Auftrittswahrscheinlichkeiten. Der Baselinetest basiert auf dem Training- und dem Test auf der kompletten Datenbank, für die Kreuzvalidierung wurde die Datenbank 100 Mal in 80% Trainings- und 20% Testmenge aufgeteilt, ohne dass ein Subjekt in beiden Mengen vorkommt.

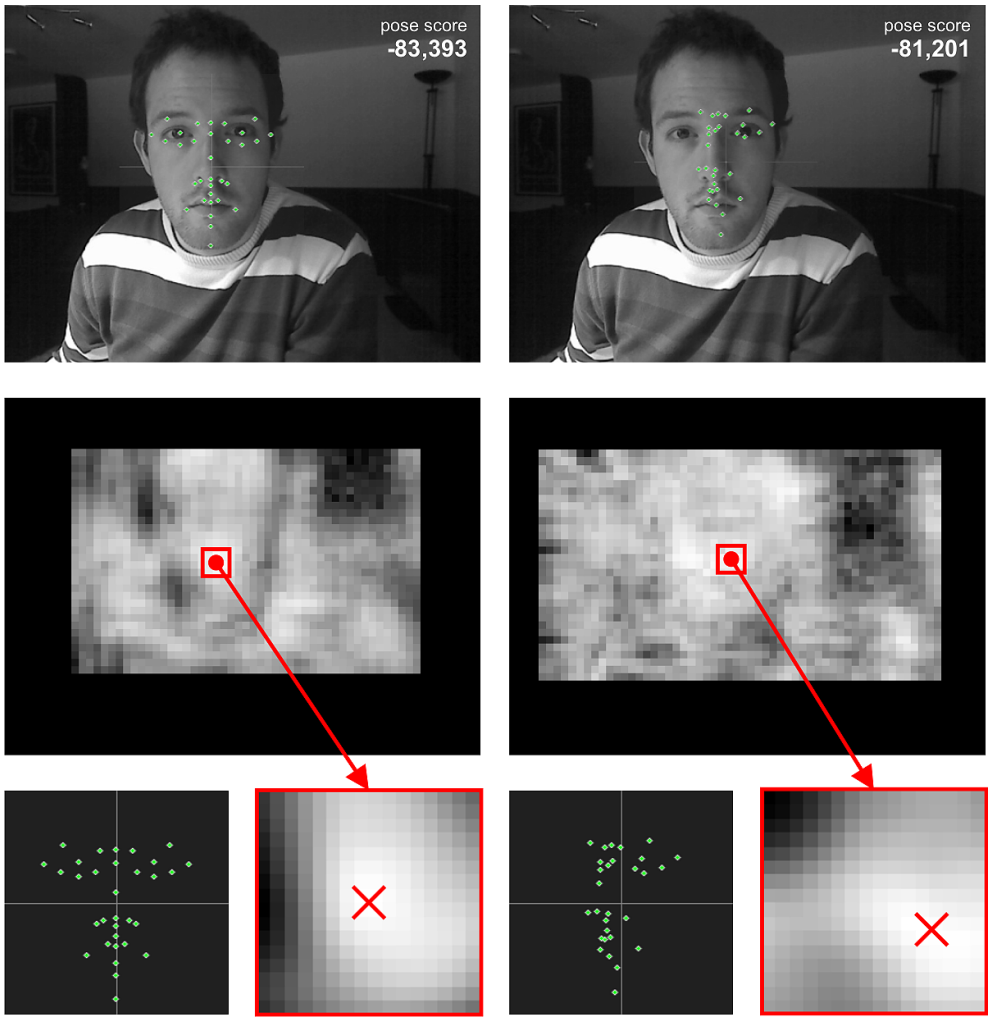
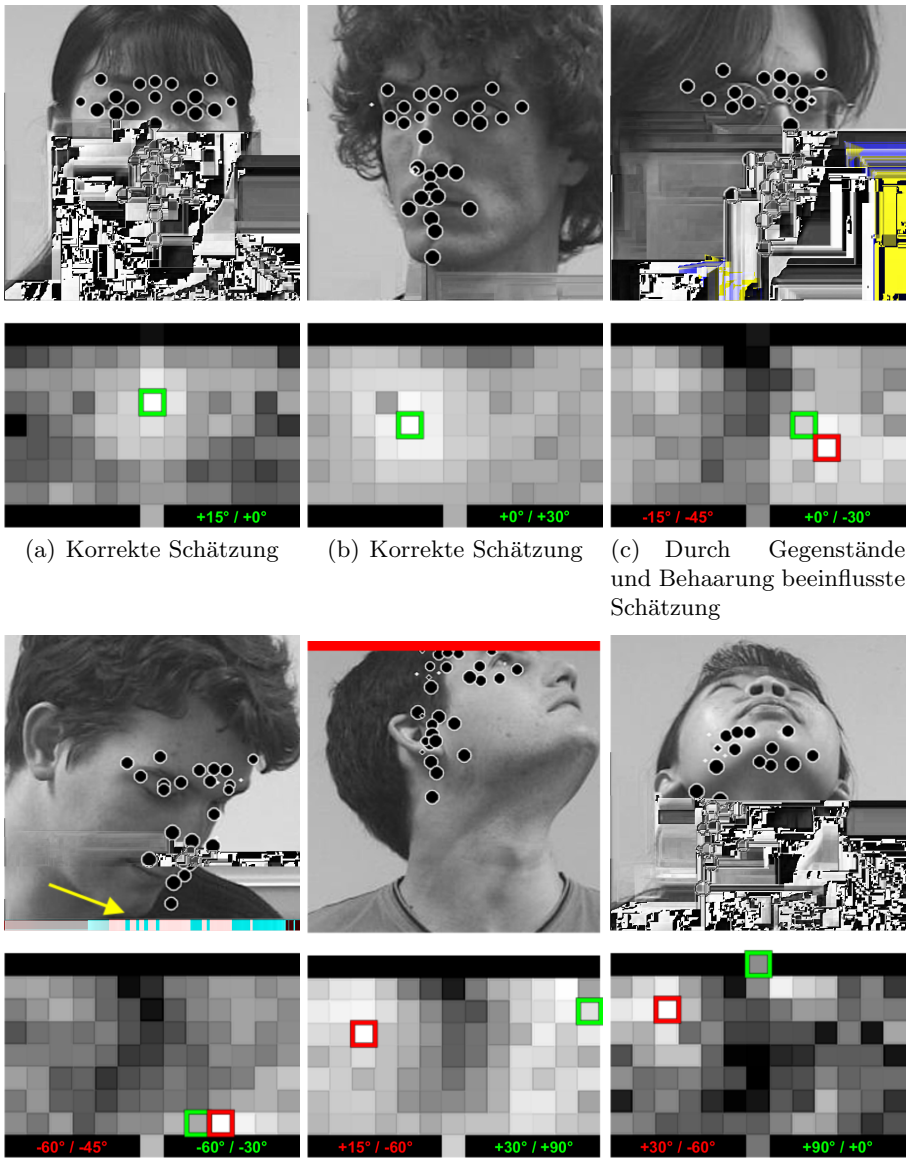


Abb. 20: Matching eines Frontal- und Halbprofil-Texturmodells auf eine Frontalansicht. Das sichelförmige Texturmodell der Halbprofilpose (rechts) wird in einer Gesichtshälfte wiedergefunden und erreicht auf diese Weise einen ähnlich hohen *pose score* $(-81,201)$ wie das Texturmodell der Frontalansichtspose (links, $-83,393$). Es sind jeweils das Suchbild (oben), der *mean model graph* des *pose models* (unten links), die *likelihood*-Karte der groben (Mitte) und feinen Rastersuche (unten rechts) abgebildet. Der *mean model graph* ist im Suchbild an der Stelle des höchsten *pose scores* eingezeichnet. In der rechten oberen Ecke des Suchbildes sind die *pose scores* angegeben. Die schwarzen Ränder in der groben *likelihood*-Karte sind die Bereiche, in denen das Texturmodell nicht geprüft werden kann, da sonst Landmarken außerhalb des Bildes liegen würden.



(a) Korrekte Schätzung

(b) Korrekte Schätzung

(c) Durch Gegenstände und Behaarung beeinflusste Schätzung

(d) Durch den Bildrand gestörte Schätzung

(e) Gesicht liegt zum Teil außerhalb des Bildrandes. *Matching* entlang der Haarlinie.

(f) *Matching* auf der Silhouette von Kinn und Hals.

Abb. 21: Testbilder der Kreuzvalidierung, die *probability maps* jeder Posenschätzung und der *mean model graph* der geschätzten Pose an $\vec{\xi}^*$. Ist die Schätzung korrekt (Abbildungen 21(a) und 21(b)), ist diese in der *probability map* mit einem grünen Quadrat markiert. Ist die Schätzung fehlerhaft (Abbildung 21(c) - 21(f)), ist mit einem roten Quadrat der geschätze Winkel und mit einem grünen Quadrat der *ground-truth*-Winkel markiert. Die Ansichten sind Ausschnitte. Ist der Bildrand zu sehen, ist er mit einem roten Balken markiert. Gelbe Pfeile deuten auf mögliche Ursachen einer fehlerhaften Schätzung hin. Die Kreise stellen die Landmarken des verwendeten *mean model graphs* dar. Die Größe eines Kreises gibt die jeweilige Auftrittswahrscheinlichkeit w_l wieder.

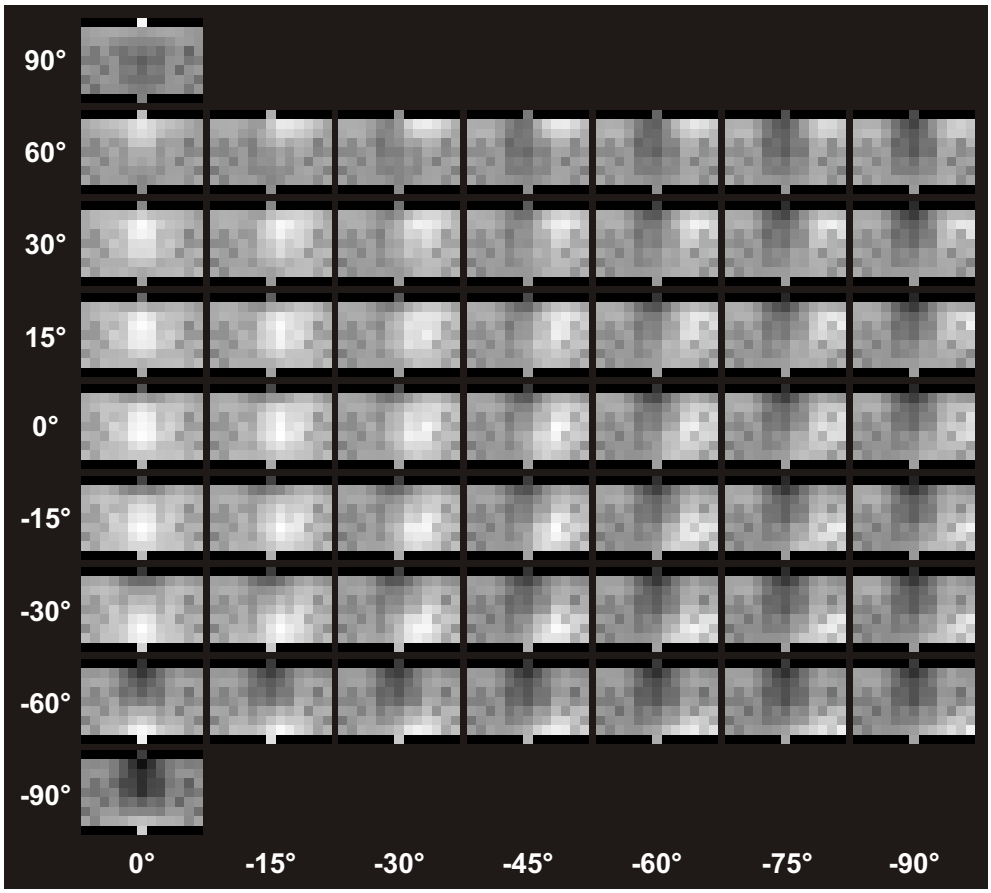


Abb. 22: Abbildung aller mittleren *pose probability maps* des halben Posenraums ($0^\circ \leq \Psi \leq -90^\circ$; X-Achse: Ψ, Y-Achse: Θ). Die Helligkeitsverteilung pro *pose probability map* gibt die Verteilung der mittleren Klassifikationen pro Kopfpose wieder. Es ist grundsätzlich zu beobachten, dass ähnlich hohe *pose scores* von benachbarten *pose models* erzeugt werden. Dieser Effekt tritt besonders in Frontalansichten auf. Für Halbprofilansichten ist zu beobachten, dass *pose models* entgegengesetzter Gierwinkel trotz des rotierten Kopfes relativ hohe *pose scores* erzielen können.

4 Diskussion

4.1 Kopflokalisation

Gesichtsdetektion Das in dieser Arbeit vorgestellte Verfahren zur Kopflokalisation ist dazu geeignet, in natürlich beleuchteten Bildern die Kopfregion zu detektieren und zu verfolgen. In künstlich beleuchteten Bildern versagt es allerdings, da die durch Beleuchtung verfälscht abgebildete Hautfarbe nicht mehr Teil des Hautfarbenmodells und deshalb die *skin probability map* fehlerhaft und uneindeutig ist. Die Wahl eines konstanten Schwellwerts ist für kontrollierte Aufnahmebedingungen möglich. Sind die Aufnahmebedingungen nicht bekannt, ist die Wahl eines niedrigen Schwellwerts ratsam, da sonst im Zweifelsfall Hautregionen nicht klassifiziert werden.

Die Detektionen des Viola/Jones-Detektors unter Verwendung von *rejection cascades* mehrerer Kopfposenansichten sind robust genug, um in einer Bildsequenz rotierte Gesichter zu detektieren. Sind die Aufnahmebedingungen unbekannt, kann ein *skin color detection approach* dazu führen, dass potentielle Gesichtsbereiche nicht als Suchbereiche registriert werden.

Kalman-Filter Trotz der Robustheit des Viola/Jones-Detektors bleiben in Bildsequenzen trotzdem häufig Detektion aus, sodass die verfolgte Gesichtsregion in solchen Situation still steht oder plötzlichen Bewegungsschwankungen ausgesetzt ist. Anstatt die Position und Rotation des Kopfes im Raum und die damit verbundene perspektivische Projektion im Bild zu berücksichtigen, wird in dieser Arbeit die Kopfregion durch unabhängige pixelbasierte Größen und lineare Änderungsraten modelliert. Für ruhige Kopfbewegungen erscheint das Bewegungsmodell geeignet, ist allerdings für arbiträre Umgebungsbedingungen nur eingeschränkt zu empfehlen.

4.2 Kopfposenerkennung

Trainingsdaten Die indirekte Messung von Winkeldaten durch Vorgabe der Blickrichtung ist generell als kritisch zu bewerten und für die Erstellung von Bilderdatenbanken von Kopfposenansichten zu vermeiden [MCT09]. Durch Schiefstellung des Kopfes variieren die tatsächlich eingenommenen Kopfposen der Subjekte der Pointing´04 Bilderdatenbank. Da die Struktur- und die lokalen Texturmerkmale der Gaborgraphen statistisch ermittelt werden, ist es generell vorteilhaft, wenn die Texturen und Positionen von Gesichtsmerkmalen in den Trainingsbildern variieren, weil das gelernte Texturmodell dadurch

invariant gegenüber leichten Deformationen und Rotationen in der Bildebene werden kann.

Gabor-basierte Texturmodelle Die hier verwendeten Texturmodelle sind dazu geeignet, Kopfposenansichten zu modellieren und mit ihnen im Rahmen eines *detector arrays* Kopfposen zu erkennen. Die Verwendung von merkmalsbasierten Auftrittswahrscheinlichkeiten stellt sich als förderlich dar, da 1.) möglicherweise verdeckte Gesichtsmerkmale beim Texturvergleich berücksichtigt und 2.) die *pose scores* von *pose models* mit unterschiedlichen Landmarken-Mengen vergleichbar gemacht werden. Die mittleren *pose probability maps* der Kreuzvalidierung weisen eindeutige *pose score*-Maxima und -Minima auf, sodass eine eindeutige Posenbestimmung möglich ist.

Für Texturmodelle von Profilansichten kann allerdings die Bildung von sichelförmigen Gesichtsmodellen beobachtet werden. Angewendet auf eine Frontalansicht, ist zu beobachten, dass diese in der linken oder rechten Gesichtshälfte „wiedergefunden“ werden. Wenn der sichelförmige Graph in diesem Fall die modellierten Texturen der gelernten Gesichtsmerkmale registrieren kann, ergibt sich ein relativ hoher *pose score* auf demselben Niveau wie von vergleichbaren Texturmodellen für Frontalansichten, da die *pose scores* auf Basis der Gewichte $\boldsymbol{\mathcal{W}}$ normalisiert werden. Aus diesem Grund sind die Messergebnisse der Klassifikationsraten für Frontalansichten niedriger als vergleichsweise für Halbprofilansichten. Die *pose scores* der Frontalansichten sind in den *pose probability maps* diffus.

Das Problem basiert auf der Auswahl der Gesichtsmerkmale des Gesichtsmodells $\boldsymbol{\mathcal{F}}$. Es ist nicht ausreichend mit Landmarken ausgestattet, um in Profilansichten genügend Texturinformationen bereitzustellen. Daher kann keine differenzierte Menge von *pose models* generiert werden. Es fehlen weitere Landmarken, die Merkmale außerhalb des Gesichts um den Kopf herum markieren. In Profilansichten, in denen das Texturmodell sichelförmig ist, würden Landmarken im konkaven Bereich der Sichel für eine präzisere Berechnung des *pose scores* sorgen, da an diesen Stellen mehrere Texturmerkmale anderer Ansichten, wie z.B. der Frontalansicht, als nicht zum Texturmodell zugehörig klassifiziert werden würden.

Es ist zu untersuchen, ob die in dieser Arbeit vorgestellten *a-priori* Kriterien zur Aufstellung von $\boldsymbol{\mathcal{F}}$ dazu geeignet sind, differenzierte Texturmodelle von Kopfposenansichten zu konstruieren.

Normalisierung Während des Trainings der Texturmodelle wird jedes Bild auf Basis des Augenaußenwinkelabstandes der Frontalansicht pro Bildsequenz normalisiert. Im Gegensatz dazu wird während der Verarbeitung eines Videos die Breite der verfolgten Gesichtsregion verwendet. Es ist zu prüfen, ob diese Methode in Kombination mit einem Viola/Jones-Detektors stabil ist. Da in dieser Arbeit unterschiedlich trainierte *rejection cascades* verwendet werden, ist nicht sichergestellt, dass die Breite der Viola/Jones-Antwort tatsächlich der der statistischen Kopfbreite entspricht.

Vergleich mit anderen Verfahren Das hier vorgestellte Verfahren eignet sich zur akkuraten Erfassung der Kopfpose. Im Vergleich mit anderen *head pose estmation*-Verfahren, für die es ebenfalls Evaluationsdaten für die Pointing´04 Datenbank gibt, kann das hier vorgestellte Verfahren den niedrigsten berichteten Nickwinkelfehler aufweisen. Die Messergebnisse reihen sich vergleichbar mit den Ergebnissen von Gourier [GMHC07] und Stiefelhagen [Sti04] ein. Das hier vorgestellte Verfahren übertrifft die in [GMHC07] ermittelten Messergebnisse zur menschlichen Einschätzung der diskreten Posen für die Pointing'04 Datenbank (vergleiche Tabelle 4).

Verfahren	Mittl. abs. Winkelfehler Nickwinkel	Gierwinkel	Mittl. Klassifikationsrate Nickwinkel	Gierwinkel
Stiefelhagen [Sti04]	9,70°	9,50°	66,3%	52%
Gourier [GMHC07]	15,90°	10,10°	43,9%	50,0%
Behrenberg	**8,87°**	**10,86°**	**62,23%**	**50,96%**
Tu (High-order SVD) [JTH07]	17,97°	12,90°	54,84%	49,25%
(PCA)	14,98°	14,11°	57,99%	55,2%
(LEA)	17,44°	15,88°	50,61%	45,16%
Voit [VNS07]	12,77°	12,30°	-	-
Zhu (PCA) [LFY+07]	35,10°	26,90°	-	-
(LDA)	26,90°	25,80°	-	-
(LPP)	22,60°	24,70°	-	-
(Local-PCA)	37,60°	24,50°	-	-
(Local-LPP)	40,20°	29,20°	-	-
(Local-LDA)	30,70°	19,10°	-	-
Menschl. Einschätzung [GMHC07]	9,40°	11,80°	59,0%	40,7%

Tab. 4: Einordnung der Messergebnisse dieser Arbeit und Vergleich mit denen anderer Kopfposenerkennungsverfahren, die für die Pointing´04 Bilderdatenbank evaluiert wurden ([MCT09], modifiziert). Sind mehrere Verfahren in einer Publikation evaluiert worden, ist in Klammern die Kurzbezeichnung des Verfahrens angegeben. Die Messergebnisse dieser Arbeit sind unter *Behrenberg* aufgeführt. Es sind jeweils der mittlere absolute Winkelfehler und die mittlere Klassifikationsrate getrennt nach Nick- und Gierwinkeln aufgeführt. Für die Verfahren von Voit [VNS07] und Zhu [LFY+07] liegen keine Messergebnisse bezüglich der Klassifikationsraten vor. Die Messergebnisse des Verfahrens von Stiefelhagen [Sti04] liegen 80% Trainingsdaten und 10% Testdaten zugrunde und sind damit mit der Evaluation des hier vorgestellten Verfahrens vergleichbar; die Messergebnisse von Gourier [GMHC07] entsprechen dem Trainingsfehler. Der absolute mittlere Nickwinkelfehler des hier vorgestellten Verfahrens entspricht dem niedrigsten Berichteten für die Pointing´04 Datenbank, die restlichen Messergebnisse stehen in Bezug auf die Genauigkeit auf der gleichen Ebene wie die Verfahren von Gourier und Stiefelhagen. Im abgesetzten Tabellenteil sind Messergebnisse zur menschlichen Einschätzung der diskreten Kopfposen der Pointing´04 Datenbank aufgeführt [GMHC07]. Das hier vorgestellte Verfahren kann diese Messwerte übertreffen.

5 Zusammenfassung und Ausblick

Kopfposenerkennung anhand statistischer Gesichtsmerkmale In dieser Arbeit wird eine Methode zur Schätzung diskreter Kopfposen vorgestellt. Dieses Verfahren ist dazu geeignet, diskrete Kopfposen anhand von auf statistischen Gesichtsmerkmalen basierenden Texturmodellen zu ermitteln. Die Erkennung findet auf normalisierten Graustufenbildern statt und basiert auf der *Gabor wavelet transformation*. Die gelernten Texturmodelle sind graphenbasiert und modellieren mithilfe eines Gaborgraphen Gesichtsmerkmale, deren Auftrittswahrscheinlichkeiten, relative Anordnungen und mittlere statistische Texturmerkmale einer Kopfposenansicht. Jedem Modell ist eine diskrete Kopfpose zugeordnet. Im Rahmen eines simplen *the winner takes all approaches* wird diejenige Pose erkannt, deren Modell die größte Ähnlichkeit mit den Bilddaten aufweist.

Mithilfe der für diese Arbeit eigens durch Landmarken annotierten Pointing'04 Datenbank werden 93 Texturmodelle von Kopfposen für Gier- und Nickwinkel für bis zu 90° (Vollprofilansichten) ermittelt. Durch eine Subjektbasierte Kreuzvalidierung mit 80% Trainings- und 20% Testmenge kann für das Verfahren eine Genauigkeit von $8,87°$ für Nick- und $10,86°$ für Gierwinkel nachgewiesen werden. Durch den geringen Winkelfehler und entsprechend hohen Klassifikationsraten konkurriert die in dieser Arbeit vorgestellte Methode mit den Messergebnissen anderen *state of the art*-Verfahren.

Die Evaluation des Verfahrens hat ergeben, dass die Verwendung von merkmalsgekoppelten Auftrittswahrscheinlichkeiten die statistische Auswertung von Bildtexturen robuster macht. Durch Okklusion beeinflusste, instabile Gesichtsmerkmale haben dadurch weniger Einfluss auf den Ähnlichkeitswert. Zwar ermöglichen die verwendeten Texturmodelle eine genaue Kopfposenschätzung, allerdings ist die Beschränkung auf die Verwendung innerer Gesichtsmerkmale kritisch. Die durch diese Beschränkung entstehenden sichelförmigen Texturmodelle von Halbprofilen besitzen in Frontalansichten vergleichbar hohe Ähnlichkeitswerte wie die Texturmodelle von Frontalansichten. Die Erweiterung möglicher Landmarken auf den gesamten Kopf und dessen Kontur erscheint geeignet, um die Genauigkeit des Verfahrens zu erhöhen.

Dynamische, kontinuierliche Kopfposenschätzung Das Verfahren wird in eine Anwendung zur dynamischen Verarbeitung von farbigen, hochauflösenden Bildsequenzen integriert. Das integrierte System wird dann zur Schätzung der kontinuierlichen Kopfpose einer Einzelperson eingesetzt. Dazu wird die Ge-

sichtsregion in einem separaten Prozess pro Bild mithilfe einer Hautfarbenerkennung und einem Viola/Jones-Detektor lokalisiert und mit einem Kalman-Filter verfolgt. Es wird gezeigt, dass unter Zuhilfenahme statistischer, anthropometrischer Messdaten eine merkmalsbasierte Bildnormalisierung über die Umrechnung der Ausmaße der verfolgten Gesichtsregion angenähert werden kann. Durch Verwendung der rechnerisch aufwändigen *Gabor wavelet transformation* ist das Verfahren nicht für die Verarbeitung von Echtzeitvideo geeignet.

Die exemplarische Anwendung zur kontinuierlichen Kopfposenerkennung ist in dieser Arbeit unidirektional aufgebaut und könnte durch Einführung von Feedback-Schleifen verbessert werden. Da das vorgestellte Verfahren zur Kopfposenerkennung eine integrierte Detektionsstufe beinhaltet, könnte die Kopflokalisation durch das Ergebnis des *detection*-Schritts der Posenerkennung verbessert werden. Ist die initiale Kopfpose bekannt, könnte der Winkelfehler reduziert und die Geschwindigkeit der Posenerkennung erhöht werden, wenn durch Berücksichtigung des temporalen Kontextes nur die benachbarten Texturmodelle der aktuellen Kopfpose getestet werden würden.

Erweiterung des Verfahrens Das in dieser Arbeit vorgestellte Verfahren zur Kopfposenerkennung könnte durch Einführung einer Skalierung, Rotation und knotenbasierten Optimierung der graphenbasierten Texturmodelle verbessert und ähnlich dem *Elastic Bunch Graph Matching* zu einem integrierten Kopfdetektions- und -posenerkennungsverfahren ausgebaut werden. Ein Nachteil des Verfahrens ist der Vergleich des gesamten Texturmodells mit dem Bild. Da die *Gabor wavelet transformation* eines Bildes eine differenzierte Betrachtung unterschiedlicher Abstraktionsebenen des Bildes erlaubt, könnte, in Anlehnung an einen *rejection cascade approach* ausgewählte, strukturelle Texturinformationen aller Texturmodelle im Rahmen eines Ablaufsplans gezielt geprüft und zur frühzeitigen Verwerfung einer Pose an einer Pixelstelle im Matchingprozess eingesetzt werden.

Literatur

[AAB+84] E.H. Adelson, C.H. Anderson, J.R. Bergen, P.J. Burt, and J.M. Ogden. Pyramid methods in image processing. *RCA engineer*, 29(6):33–41, 1984.

[Bey94] David J. Beymer. Face recognition under varying pose. *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, pages 756–761, 1994.

[Bis05] Christopher M. Bishop. *Neural networks for pattern recognition*. Clarendon Press, Oxford, 2005.

[BK08] G. Bradski and A. Kaehler. *Learning OpenCV: Computer vision with the OpenCV library*. O'Reilly Media, Inc., 2008.

[BL85] P. Bergeron and P. Lachapelle. Controlling facial expressions and body movements in the computer-generated animated short tony de peltrie. In *SIGGRAPH 85 Advanced Computer Animation seminar notes*, 1985.

[Boa] OpenMP Architecture Review Board. C++ application program interface. *Available at: http://www.openmp.org*.

[Bou04] N. Bourbaki. *Theory of sets*. University of British Columbia Press, 2004.

[Bru09] R. Brunelli. *Template Matching Techniques in Computer Vision: Theory and Practice*. Wiley-Blackwell, 2009.

[BT95] I.T.U.R.R. BT. 601-5:studio encoding parameters of digital television for standard 4: 3 and wide-screen 16: 9 aspect ratios.. *ITU, Geneva, Switzerland*, 1995.

[Bur83] P.J. Burt. Fast algorithms for estimating local image properties. *Computer Vision, Graphics, and Image Processing*, 21(3):368–382, 1983.

[BW01] G. Bishop and G. Welch. An introduction to the kalman filter. *SIGGRAPH 2001 Course Notes*, 2001.

[CCL04] F. Chang, C.J. Chen, and C.J. Lu. A linear-time component-labeling algorithm using contour tracing technique. *Computer Vision and Image Understanding*, 93(2):206–220, 2004.

[Cha00] R. Chandra. *Parallel programming in OpenMP*. Morgan Kaufmann, 2000.

[CHLH08] Matthew B. Blaschko Christoph H. Lampert and Thomas Hofmann. Beyond sliding windows: Object localization by efficient subwindow search. In *Proc. of CVPR*, volume 1, page 3, 2008.

[Cor] Intel Corporation. The opencv open source computer vision library. Available: http://opencvlibrary.sourceforge.net.

[CPPP98] Michael Oren Constantine P. Papageorgiou and Tomaso Poggio. A general framework for object detection. In *Proceedings of the Sixth International Conference on Computer Vision*, pages 555–562. IEEE Computer Society, 1998.

[Dau85] J.G. Daugman. Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters. *Journal of the Optical Society of America A*, 2(7):1160–1169, 1985.

[Eme00] N.J. Emery. The eyes have it: the neuroethology, function and evolution of social gaze. *Neuroscience and Biobehavioral Reviews*, 24(6):581–604, March 2000.

[ESS04] Kai Nickel Edgar Seemann and Rainer Stiefelhagen. Head pose estimation using stereo vision for human-robot interaction. In *Sixth IEEE International Conference on Automatic Face and Gesture Recognition*, pages 626–631, 2004.

[Far94] L.G. Farkas. *Anthropometry of the Head and Face*. Raven Press, New York, 1994.

[FBH97] H.S. Fairman, M.H. Brill, and H. Hemmendinger. How the cie 1931 color-matching functions were derived from wright-guild data. *Color Research & Application*, 22(1):11–23, 1997.

[FS97] Y. Freund and R.E. Schapire. A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of computer and system sciences*, 55(1):119–139, 1997.

[FS01] Yoav Freund and Robert E. Schapire. Experiments with a new boosting algorithm. *Proceedings of the 13th International Conference on Machine Learning*, pages 148–156, 2001.

[Gab46] D. Gabor. Theory of communication. *J. Inst. Electr. Eng.*, 93:429–457, 1946.

[GC94] A. Gee and R. Cipolla. Determining the gaze of faces in images. *Image and Vision Computing*, 12(10):639–647, 1994.

[GC96] A. Gee and R. Cipolla. Fast visual tracking by temporal consensus. *Image and Vision Computing*, 14(2):105–114, 1996.

[GCPC95] H.P. Graf, T. Chen, E. Petajan, and E. Cosatto. Locating faces and facial parts. *Proc. First Intl Workshop Automatic Face and Gesture Recognition*, pages 41–46, 1995.

[GHC04] N. Gourier, D. Hall, and J.L. Crowley. Estimating face orientation from robust detection of salient facial structures. In *FG Net Workshop on Visual Observation of Deictic Gestures (POINTING)*, 2004.

[GMHC07] N. Gourier, J. Maisonnasse, D. Hall, and J.L. Crowley. Head pose estimation on low resolution images. *Lecture Notes in Computer Science*, 4122:270, 2007.

[GW09] Manuel Günther and Rolf P. Würtz. Face detection and recognition using maximum likelihood classifiers on gabor graphs. *International Journal of Pattern Recognition and Artificial Intelligence*, 23(3):433–461, May 2009.

[GW10] Manuel Günther and Rolf P. Würtz. Reconstruction of images from graphs labeled with Gabor jets. 2010. In preparation.

[Haa10] A. Haar. Zur theorie der orthogonalen funktionensysteme. *Mathematische Annalen*, 69(3):331–371, 1910.

[HARK96] Shumeet Baluja Henry A. Rowley and Takeo Kanade. Human face detection in visual scenes. *Advances in Neural Information Processing Systems*, pages 875–881, 1996.

[Her64] E. Hering. *Outlines of a Theory of the Light Sense*. Harvard University Press Cambridge, MA, 1964.

[HSW98] J. Huang, X. Shao, and H. Wechsler. Face pose discrimination using support vector machines (svm). In *International Conference on Pattern Recognition*, volume 14, pages 154–156. Citeseer, 1998.

[JP87] Judson P. Jones and Larry A. Palmer. An evaluation of the two-dimensional gabor filter model of simple receptive fields in cat striate cortex. *Journal of Neurophysiology*, 58(6):1233–1258, 1987.

[JR02] M.J. Jones and J.M. Rehg. Statistical color models with application to skin detection. *International Journal of Computer Vision*, 46(1):81–96, 2002.

[JTH07] Yuxiao Hu Jilin Tu, Yun Fu and Thomas Huang. Evaluation of head pose estimation for studio data. *Lecture Notes in Computer Science*, 4122:281, 2007.

[JV03] Michael J. Jones and Paul Viola. Fast multi-view face detection. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2003.

[JY02] Q. Ji and X. Yang. Real-time eye, gaze, and face pose tracking for monitoring driver vigilance. *Real-Time Imaging*, 8(5):357–377, 2002.

[Kal60] R.E. Kalman. A new approach to linear filtering and prediction problems. *Journal of basic Engineering*, 82(1):35–45, 1960.

[KMB07] P. Kakumanu, S. Makrogiannis, and N. Bourbakis. A survey of skin-color modeling and detection methods. *Pattern Recognition*, 40(3):1106–1122, 2007.

[KP92] J.M. Kasson and W. Plouffe. An analysis of selected computer interchange color spaces. *ACM Transactions on Graphics (TOG)*, 11(4):373–405, 1992.

[KW94] R.R. Knipling and J.S. Wang. Crashes and fatalities related to driver drowsiness/fatigue. *Research note, Washington DC Department of Transportation, National Highway Traffic Safety Administration*, 1994.

[LFY+07] Z. Li, Y. Fu, J. Yuan, T.S. Huang, and Y. Wu. Query driven localized linear discriminant models for head pose estimation. In *Proc. IEEE Intl. Conf. Multimedia and Expo*, pages 1810–1813, 2007.

[LKP03] R. Lienhart, A. Kuranov, and V. Pisarevsky. Empirical analysis of detection cascades of boosted classifiers for rapid object detection. *Lecture Notes in Computer Science*, pages 297–304, 2003.

[LM02] Rainer Lienhart and Jochen Maydt. An extended set of haar-like features for rapid object detection. In *IEEE ICIP*, volume 1, pages 900–903, 2002.

[LWvdM97] Norbert Krüger Laurenz Wiskott, Jean-Marc Fellous and Christoph von der Malsburg. Face recognition by elastic bunch graph matching. *IEEE Transactions on pattern analysis and machine intelligence*, 19(7):775–779, 1997.

[M+89] S.G. Mallat et al. A theory for multiresolution signal decomposition: The wavelet representation. *IEEE transactions on pattern analysis and machine intelligence*, 11(7):674–693, 1989.

[Mar71] A.A. Markov. Extension of the limit theorems of probability theory to a sum of variables connected in a chain. *Dynamic Probabilistic Systems*, 1, 1971.

[MCT09] E. Murphy-Chutorian and M.M. Trivedi. Head pose estimation in computer vision: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pages 607–626, 2009.

[Men00] Alberto Menache. *Understanding motion capture for computer animation and video games*. Morgan Kaufmann Publishers, 2000.

[Men05] Fabio Meneghini. *Clinical Facial Analysis: Elements, Principles and Techniques*. Springer, 1 edition, 3 2005.

[MG98] S.J. McKenna and S. Gong. Real-time face pose estimation. *Real-Time Imaging*, 4(5):333–347, 1998.

[MKMW07] A. H. Tewes A. Schäfer M. K. Müller, A. Heinrichs and R. P. Würtz. Similarity rank correlation for face recognition under unenrolled pose. *Lecture Notes in Computer Science*, 4642:6776, 2007.

[MW09] M.K. Müller and R.P. Würtz. Learning from examples to generalize over pose and illumination. 2009.

[NF96] S. Niyogi and W.T. Freeman. Example-based head tracking. In *Proceedings of the 2nd International Conference on Automatic Face and Gesture Recognition (FG'96)*, page 374. IEEE Computer Society Washington, DC, USA, 1996.

[NKvdM97] Michael Pötzsch Norbert Krüger and Christoph von der Malsburg. Determination of face position and pose with a learned representation based on labelled graphs. *Image and Vision Computing*, 15(8):665–673, 1997.

[Nor79] D. Normen. Din 6174: Farbmetrische bestimmung von farbabständen bei körperfarben nach der cie-lab-formel. *Normenausschuß Farbe (FNF) im DIN Deutsches Institut für Normung e.V., Beuth, Berlin Januar*, 1979.

[PH07] Simon Perreault and Patrick Hébert. Median filtering in constant time. *IEEE Transactions on Image Processing*, 16(9):2389–2394, 2007.

[QHTK02] Mitsuhiko Meguro Quan Huynh-Thu and Masahide Kaneko. Skin-color-based image segmentation and its application in face detection. In *IAPR Workshop on Machine Vision Applications*, pages 48–51, 2002.

[SG31] T. Smith and J. Guild. The cie colorimetric standards and their use. *Transactions of the Optical Society*, 33:73–134, 1931.

[Sti04] Rainer Stiefelhagen. Estimating head pose with neural networks-results on the pointing04 icpr workshop evaluation data. In *Proc. Pointing 2004 Workshop: Visual Observation of Deictic Gestures*, 2004.

[SYW02] R. Stiefelhagen, J. Yang, and A. Waibel. Modeling focus of attention for meeting indexing based on multiple cues. *IEEE Transactions on Neural Networks*, 13(4):928–938, 2002.

[TEHF89] Worthy N. Martin Kelly C. Reichert Thomas E. Hutchinson, K. Preston White Jr. and Lisa A. Frey. Human-computer interaction using eye-gaze input. *IEEE Transactions on systems, man and cybernetics*, 19(6):1527–1534, 1989.

[TK91] Carlo Tomasi and Takeo Kanade. Detection and tracking of point features. Technical report, Carnegie Mellon University, 1991.

[Tuk77] John Wilder Tukey. *Exploratory data analysis*. Addison-Wesley Menlo Park, CA., 1977.

[VH67] H. Von Helmholtz. *Handbuch der physiologischen Optik*. Voss, 1867.

[VJ01] P. Viola and M. Jones. Rapid object detection using a boosted cascade of simple features. In *Proc. CVPR*, volume 1, pages 511–518, 2001.

[vL27] Felix von Luschan. *Völker, Rassen, Sprachen - Anthropologische Beobachtungen*. 1927.

[VNS07] M. Voit, K. Nickel, and R. Stiefelhagen. Neural network-based head pose estimation and multi-view fusion. *Lecture Notes in Computer Science*, 4122:291, 2007.

[VVA03] Vassili Sazonov Vladimir Vezhnevets and Alla Andreeva. A survey on pixel-based skin color detection techniques. pages 85–92, 2003.

[Wür95] Rolf P. Würtz. *Multilayer dynamic link networks for establishing image point correspondences and visual object recognition*. PhD thesis, 1995.

[YKA02] M.H. Yang, D.J. Kriegman, and N. Ahuja. Detecting faces in images: A survey. *IEEE Transactions on Pattern analysis and Machine intelligence*, pages 34–58, 2002.

[You93] J.W. Young. Head and face anthropometry of adult us civilians. 1993.

[YZ02] R. Yang and Z. Zhang. Model-based head pose tracking with stereovision. In *Proc. Fifth IEEE International Conference on Automatic Face and Gesture Recognition*, pages 255–260, 2002.

[Z+76] S.W. Zucker et al. Region growing: Childhood and adolescence. *Computer graphics and image processing*, 5(3):382–399, 1976.

[ZCPR03] W. Zhao, R. Chellappa, PJ Phillips, and A. Rosenfeld. Face recognition: A literature survey. *Acm Computing Surveys (CSUR)*, 35(4):399–458, 2003.

A Farbraumtransformationen

RGB $\mapsto$ Graustufen Die Intensität I eines RGB Tupels wird durch Gewichtung [BT95] seiner Komponenten berechnet:

$$I = 0,299 \cdot R + 0,587 \cdot G + 0,114 \cdot B \tag{21}$$

RGB $\mapsto$ CIE-L*a*b Für die Transformation $RGB \mapsto LAB$ müssen die RGB-Komponenten normalisiert und zuvor wie folgt [FBH97] in das CIE-XYZ Normvalenzsystem [SG31] überführt werden:

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} 0,412453 & 0,35758 & 0,180423 \\ 0,212671 & 0,715160 & 0,072169 \\ 0,019334 & 0,119193 & 0,950227 \end{bmatrix} \cdot \begin{bmatrix} R/255 \\ G/255 \\ B/255 \end{bmatrix} \tag{22}$$

LAB Farbtupel haben den Wertebereich $0 \leq L \leq +100$, $-128 \leq A \leq +127$ und $-128 \leq B \leq +127$. Die Transformation $XYZ \mapsto LAB$ ist nicht linear:

$$L = \begin{cases} 116 \cdot \sqrt[3]{Y} - 16 & \text{für } Y > 0,008856 \\ 903,3 \cdot Y & \text{sonst} \end{cases} \tag{23}$$

$$A = 500 \cdot \left(f\left(\frac{X}{0,950456}\right) - f(Y) \right)$$

$$B = 200 \cdot \left(f(Y) - f\left(\frac{Z}{1,088754}\right) \right)$$

für:

$$f(\alpha) = \begin{cases} \sqrt[3]{\alpha} & \text{für } \alpha > 0,008856 \\ 7,787 \cdot \alpha + \frac{16}{116} & \text{sonst} \end{cases}$$

Für die rechnerkompatible 8-Bit Darstellung der LAB Farbe müssen die Komponenten linear transformiert und diskretisiert werden:

$$L' \leftarrow \left\lfloor \frac{255 \cdot L}{100} \right\rfloor \quad A' \leftarrow \lfloor A + 128 \rfloor \quad B' \leftarrow \lfloor B + 128 \rfloor \tag{24}$$

B Daten anthropomorpher Gesichtsmessungen

Die in Tabelle 2 zusammengefassten Messwerte beziehen sich auf die Datenblätter der Messungen 2 (Kopfbreite, *head breadth*; siehe Tabelle 5) und 6 (Abstand der Augenaußenwinkel, *biectocanthus breadth*; siehe Tabelle 6) aus der Studie von Young [You93]. Die Datenblätter sind hier in modifizierter Form abgebildet.

HEAD BREADTH

FEMALE			MALE		
SUMMARY STATISTICS			SUMMARY STATISTICS		
MILLIMETERS		INCHES	MILLIMETERS		INCHES
145.76	MEAN	5.74	152.39	MEAN	6.ØØ
Ø.43	STD ERROR(MEAN)	Ø.Ø2	Ø.41	STD ERROR(MEAN)	Ø.Ø2
5.57	STD DEVIATION	Ø.22	5.34	STD DEVIATION	Ø.21
128.Ø2	MINIMUM	5.Ø4	138.94	MINIMUM	5.47
161.Ø4	MAXIMUM	6.34	166.12	MAXIMUM	6.54
COEFF. OF VARIATION(%)		3.8Ø	COEFF. OF VARIATION(%)		3.5Ø
SYMMETRY		Ø.Ø2	SYMMETRY		-Ø.Ø1
KURTOSIS		3.25	KURTOSIS		2.48
NUMBER OF SUBJECTS		169	NUMBER OF SUBJECTS		167

Tab. 5: Auflistung statistischer Messwerte zur Untersuchung der Kopfbreite US-amerikanischer Personen [You93]. Bezieht sich auf die maximale Ausdehnung zwischen der linken und rechten Seite des Kopfes.

BIECTOCANTHUS BREADTH

FEMALE			MALE		
SUMMARY STATISTICS			SUMMARY STATISTICS		
CENTIMETERS		INCHES	CENTIMETERS		INCHES
88.53	MEAN	3.49	91.31	MEAN	3.6Ø
Ø.41	STD ERROR(MEAN)	Ø.Ø2	Ø.5Ø	STD ERROR(MEAN)	Ø.Ø2
4.78	STD DEVIATION	Ø.19	4.99	STD DEVIATION	Ø.2Ø
76.96	MINIMUM	3.Ø3	82.Ø4	MINIMUM	3.23
1Ø3.89	MAXIMUM	4.Ø9	1Ø7.95	MAXIMUM	4.25
COEFF. OF VARIATION(%)		5.41	COEFF. OF VARIATION(%)		5.43
SYMMETRY		Ø.44	SYMMETRY		Ø.57
KURTOSIS		3.Ø1	KURTOSIS		3.41
NUMBER OF SUBJECTS		134	NUMBER OF SUBJECTS		1Ø1

Tab. 6: Auflistung statistischer Messwerte zur Untersuchung des Abstandes der Augenaußenwinkel US-amerikanischer Personen [You93]. Bezieht sich auf die Distanz zwischen den lateralen (äußeren) Augenwinkelfalten.

Erklärung

Ich erkläre, dass das Thema dieser Arbeit nicht identisch ist mit dem Thema einer von mir bereits für ein anderes Examen eingereichten Arbeit.

Ich erkläre weiterhin, dass ich die Arbeit nicht bereits an einer anderen Hochschule zur Erlangung eines akademischen Grades eingereicht habe.

Ich versichere, dass ich die Arbeit selbstständig verfasst habe und keine anderen als die angegebenen Quellen benutzt habe. Die Stellen der Arbeit, die anderen Werken dem Wortlaut oder dem Sinn nach entnommen sind, habe ich unter Angabe der Quellen der Entlehnung kenntlich gemacht. Dies gilt sinngemäß auch für gelieferte Zeichnungen, Skizzen und bildliche Darstellungen und dergleichen.

Datum	Unterschrift