Schnelle Verfahren zur Objektregistrierung in der Bildverarbeitung am Beispiel der Gesichtsstabilisierung

Jannis Bloemendal
Fachhochschule Köln, Institut für Informatik

22. Februar 2009

Abstract

This paper describes fast image registration methods applied to face stabilisation. For registration accuracy comparison a new measure, which represents the visual perseption, is presented. To evaluate the methods an extensive test set of image sequences is created. An optimization of an existing algorithm and a new method based on Logarithmic Search and Normalized Cross Correlation is provided. The new method accomplish good results and is capable to create superresolution images of an image sequence which improve face recognition.

1 Einleitung

Bilder sind heutzutage ein wichtiges Beweismittel für kriminaltechnische Recherchen und Ermittlungen. Sie werden z.B. durch Videoüberwachungssysteme in vielen Situationen des alltäglichen Lebens erfasst und bieten umfangreiche Informationen.

Jedoch birgt die Erkennung, Aufzeichnung und Wiedererkennung von Gesichtern in Bildquellen Hindernisse, die durch optische Störeinflüsse wie Beleuchtung und Begrenzungen durch die Digitalisierung gegeben sind. CCD-Sensoren sind durch Auflösung, Wertebereich sowie Abtastrate limitiert und rufen hierdurch Artefakte wie Blurring, Rauschen und Aliasing[1] hervor, die nicht dem realen Bild entsprechen. Die Idee besteht darin, aus verschiedenen Aufnahmen eines Gesichts, die aus einer Sequenz stammen, mehr Informationen zu gewinnen, als ein Einzelbild der Sequenz bereit hält und somit die Wiedererkennung zu verbessern. Dieses wird als Superresolution oder Noise Reduction bezeichnet [1][2].

Das Ziel ist zwischen einem Referezbild und jedem anderen Bild der Sequenz, (die wir Objektbilder nennen wollen), eine Transformation zu finden, so dass Referenzbild und registriertes Objektbild sich maximal ähnlich sind. Die registrierten Bilder werden dann skaliert und gemittelt um die Auflösung zu erhöhen und Artefakte zu unterdrücken.

Der Arbeit liegt ein Matlab-Implementierung eines echtzeitfähigen Algorithmus nach Kourogi [3] [4] zugrunde, der optimiert wurde und einem neuen Algorithmus von Wolfgang Konen [5] gegenübergestellt wird. Hierfür wird ein neues Vergleichsmaß, das der visuellen Wahrnehmung entspricht vorgestellt. Die Bildsequenzen wurden mit dem FaceSnapRecorder der Firma CrossMatch Technologies Inc. erstellt. Dieser erkennt in Bildquellen Gesichter und speichert diese im Gif-Format mit einem zusätzlichen Kommentar, in dem die Bounding Box (Gesichtsregion) hinterlegt ist, ab.

2 Testaufnahmen

Das Ziel der Arbeit ist mehrere Bilder einer Sequenz auf ein Referenzbild zu registieren. Die Bilder sollen hierfür relativ ähnlich sein. Das bedeutet, dass eine perspektivische Bewegung des Kopfes nur bis zu einem gewissen Grad abgebildet werden kann, da der Kopf bzw. das Gesicht nicht durch ein Verfahren erfasst wird, welches ein dreidimensionales Modell des Gesichtes berücksichtigt. Die primären Transformationen entstehen durch leichte Bewegung des Kopfes:

- Die Person bewegt sich auf die Kamera zu,

[1] Eine zu geringe Auflösung von CCD-Sensoren kann dazu führen, dass für feine Strukturen das Nyquist-Shannon-Theorem in der Abtastung des Signals nicht eingehalten werden kann, welches als Artefakt im Bild sichtbar wird.

oder von ihr weg (Skalierung)

- Leicht translatorische Bewegungen, die zum Beispiel durch das Laufen einer Person entstehen können
- Leichte Neigung des Kopfes nach links oder rechts (Rotation)
- Leichte Bewegung des Kopfes nach links oder rechts bzw. oben und unten (perspektivisch)

Für die Evaluation der nachfolgenden Methoden wurden 68 Sequenzen aufgenommen, in denen Bewegungen, Beleuchtung und Größe variiert wurden um eine weites Spektrum an Szenarien abzudecken. Hierbei wurden drei Hauptgruppen kategorisiert:

- Aufnahmen mit Variation der Beleuchtung
- Aufnahmen mit Variation des Rauschanteils im Bild
- Aufnahmen unter normalen Bedingungen

3 Methoden

Die nachfolgenden Methoden wurden für die Registrierung der Bilder verwendet.

3.1 Vergleichsmaß

Um die Güte des Matches der Registrierung zu messen, benötigt man ein Fehlermaß, nachdem man optimieren kann. Die Bounding Box wird hierfür in $n \times n$ gleich große Templates aufgeteilt und für alle n^2 zueinander gehörenden Templatepaare A_{ij},B_{ij} der Korrelationskoeffizenten berechnet und gemittelt (Abb. 1). Die Methode ermöglicht eine relative Abstufung der Qualität, sowie den absoluten Vergleich der Güte der Registrierung verschiedener Sequenzen und Registrierungsverfahren. Globale Hellikeitsunterschiede sowie Rauschen werden hierbei gut kompensiert.

$$mean_{C_L} = \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} C_L(A_{ij}, B_{ij})}{n^2} \quad (1)$$

Die folgenden Fehlerwerte wurden mit $n = 5$ Templates berechnet (vgl. Abschnitt 4). Für die Untersuchung der Verfahren wurde aus den Testsequenzen eine Auswahl getroffen, so dass aus jeder

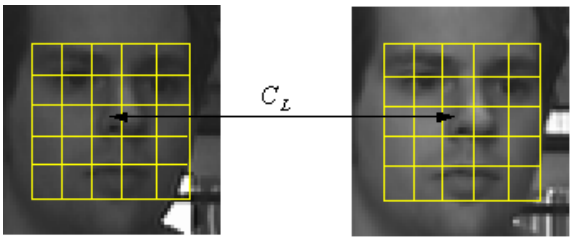


Abbildung 1: $mean_{C_L}$ Einteilung der Bilder in Templates: Die Bounding Box des Referenzbildes wird in gleich große Bereiche eingeteilt. Die Einteilung wird für Referenz- und registriertes Objektbild verwendet um die normierten Korrelationskoeffizienten für die einzelnen Ausschnitte (Templates) zu berechnen und über diese zu mitteln.

Bildkategorie je fünf Bildpaare vorhanden sind die den folgenden Fehlerwerten, (durch Einsatz der Methode von Kourogi ohne die Optimierungen), entsprechen:

1. $0.8 < mean_{C_L} \leq 1$: guter Match
2. $0.4 < mean_{C_L} \leq 0.8$: mittel schlechter Match
3. $mean_{C_L} \leq 0.4$: schlechter Match

Die Einteilung soll eine differenzierte Betrachtung der Auswirkung der Methoden ermöglichen.

3.2 Optimierung Kourogi

Das Verfahren von Kourogi[3] ist für eine andere Problemstellung konzipiert als das Vorhaben dieser Arbeit. Für die Anpassung wird für die Registrierung die affine Transformation T mit den Freiheitsgraden a_{ij} um die Scherung verringert, da diese in den Aufnahmen naturgemäß nicht vorkommt.

$$\begin{pmatrix} x' \\ y' \\ 1 \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ -a_{12} & a_{11} & a_{23} \\ 0 & 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} x \\ y \\ 1 \end{pmatrix} \quad (2)$$

Eine initiale Schätzung der Transformation T_i^{-1} (Target-to-Source Resampling) wird durch die Größe und Lage der Bounding Boxen, (gegeben mit Höhe $s_i^{(x)}$, Breite $s_i^{(y)}$ und Zentrum (x_i, y_i)), des Referenz- und Objektbildes bestimmt.

$$T_i^{-1} = \begin{pmatrix} s_i^{(x)}/s_1^{(x)} & 0 & x_1 - x_i \\ 0 & s_i^{(x)}/s_1^{(x)} & y_1 - y_i \\ 0 & & 1 \end{pmatrix} \quad (3)$$

Die Annahme der konstanten Lichtquelle, auf der die Idee des optischen Flusses und somit das Verfahren von Kourogi [3] beruht, ist in den Aufnahmen nicht immer erfüllt. Menschen laufen in observierten Bereichen unter Lichtquellen hindurch und es gibt Lichteinfall aus mehreren Richtungen. Das Verfahren von Kourgi führt hierdurch zu keiner geeigneten Transformation. Um diesen Umstand zu kompensieren werden die Bilder durch die Subtraktion ihrer tiefpassgefilterten Version angeglichen. Hierfür wird ein rekursiver IIR-Filter [6] verwendet. Zunächst wird für eine Dimension auf den Pixeln $k = 1, ..., N-1$ eine In-Place-Nach-Rechts-Rekursion

$$I_k = (I_k + aI_{k-1})\frac{1}{1+a}, \qquad k = 1, ..., N-1 \quad (4)$$

und auf dem Ergebnis eine In-Place-Nach-Links-Rekursion

$$I_k = (I_k + aI_{k+1})\frac{1}{1+a}, \qquad k = N-2, ..., 0 \quad (5)$$

mit dem Glättungsparameter a (und danach für die andere Dimension) durchführt.

Das Verfahren von Kourogi berücksichtigt keine Ausreißer, die eine falsche Zuordnung von Region des gleichen Grauwertes entsprechen. Diese Falsch-Zuordnungen können durch große Gradienten entstehen, die die Iteration und somit die Compensated Motion in eine falsche Richtung lenkt. Um dieses zu vermeiden, wurde eine Betrachtung der Pseudo Motion Vektoren (u_p, v_p) durchgeführt, die den Grauwertdifferenz-Test passieren, dessen geschätzte Verschiebung

$$\left| \begin{pmatrix} -I_t^{(c)}/I_x \\ -I_t^{(c)}/I_y \end{pmatrix} \right| < T_{maxmotion} \quad (6)$$

jedoch von der Gesamtheit abweicht. Durch eine Untersuchung der Beträge der Vektoren mehrerer Testsets wurde anhand eines Boxplots (Abb. 2) eine Schranke gefunden, die einen maximalen Betrag $T_{maxmotion} = 15$ für die Verschiebungsvektoren vorgibt und das Verfahren stabilisiert.

Die Kombination der Optimierungen mit $a = 4.5$ für die Beleuchtungsanpassung, der Begrenzung der Ausreißer durch $T_{maxmotion} = 15$ und der initialen Schätzung der Transformation hat zu guten Resultaten geführt.

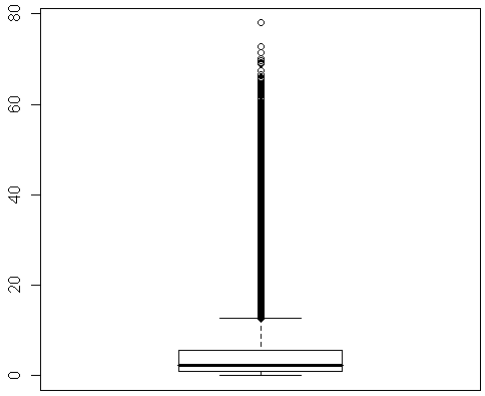


Abbildung 2: Verteilung der Verschiebungsvektorbeträge für eine Auswahl von je sechs Bildpaare aus den guten, mittel-schlechten und schlechten Matches

3.3 Neuer Algorithmus

Die Idee des neuen Algorithmus von Wolfgang Konen [5] besteht darin, für je zwei gleich große Templates A und B des Referenz- und Objektbildes eine Entsprechung zu finden, sodass der Korrelationskoeffizient C_L maximal wird. Durch eine initiale Schätzung der Position eines Templates, lässt sich im Umfeld der Schätzung durch translatorische Verschiebungen ein optimaler Match anhand des maximalen Korrelationskoeffizienten ermitteln. Für die Suche des Matches wird das Logarithmic Search Verfahren [7] verwendet. Sei durch $r_s = 2s + 1$ eine (ungerade) Templategröße definiert. Für eine minimale Anzahl an Landmarken[2] $accept_{min}$ und die Begrenzung von Ausreißern durch u_{thresh}, ist die Vorgehensweise des Algorithmus:

1. Initiale Schätzung von (u_c, v_c)
2. Verteile $n \cdot n$ Landmarken L_l gleichmäßig auf dem Referenzbild
3. Für jede Landmarke $L_l = (x, y)$ finde im Umfeld von (u_c, v_c) mittels Log-Search die Position $(u_p, v_p) = (u_c + u_{adj}, v_c + v_{adj})$ im Objekt-

[2]Position in dem Referenzbild, für die eine Entsprechung im Objektbild gesucht wird

bild I_i für die gilt:

$$\begin{aligned} C_L(A,B) = max, \qquad \text{mit} \\ A = \{I_1(m,n) \mid x - s \leq m \leq x + s \wedge \\ y - s \leq n \leq y + s\} \\ B = \{I_i(m,n) \mid x + u_p - s \leq m \leq x + u_p + s \wedge \\ y + v_p - s \leq n \leq y + v_p + s\} \end{aligned} \tag{7}$$

und der Korrelations-Koeffizient ist größer als eine Schranke

$$C_L(A,B) \geq c_{min} \tag{8}$$

4. Wenn weniger als $accept_{min}\%$ der $n \cdot n$ Landmarken den c_{min}-Test bestehen, so nehme die $accept_{min}\%$ Landmarken mit dem größten $C_L(A,B)$.

5. Berechne mittels der Methode der Kleinsten Quadrate die Transformation und hierdurch die Compensated Motion (u_c, v_c) neu.

6. Wenn der euklidische Abstand

$$\sqrt{(u_p - u_c)^2 + (v_p - v_c)^2} > u_{thresh} \tag{9}$$

zwischen der gefundenen Position (u_p, v_p) und der Schätzung (u_c, v_c) einer Landmarke größer ist als u_{thresh}, dann verwerfe sie.

7. Wenn weniger als $accept_{min}\%$ der $n \cdot n$ Landmarken den u_{thresh}-Test bestehen, so nehme die $accept_{min}$ Landmarken mit dem kleinsten euklidischen Abstand.

8. Berechne mittels der Methode der Kleinsten Quadrate die Transformation

Das Verfahren ist durch die Tests und die normalisierte Kreuzkorrelation relativ robust. Die Tests verhindern, richtig parametrisiert, Ausreißer, und der Kontrollparameter $accept_{min}$ verhindert, dass zu wenig Landmarken in die Berechnung der Transformation einfließen.

Als Parameter wurden die von Wolfgang Konen ermittelten Werte eingesetzt: $r_s = 15, accept_{min} = 0.20, u_{thresh} = 0.85, c_{min} = 2.5$ und eine Suchkreuzgröße von 4 für das Logarithmic Search Verfahren.

Abbildung 3: Künstlicher Fehler: Durchschnittsbild des Referenz- und registrierten Objektbildes, in das nachträglich von links nach rechts folgende Fehler eingerechnet wurden: w - $0; 0.3; 0.6; 0.9$

4 Ergebnisse

Durch Einführung eines künstlichen Registierungsfehlers mit dem Fehlergewicht $w \in [0,1]$

$$\begin{aligned} A_i' &= A_i \cdot \begin{pmatrix} \cos(\alpha) + s & -\sin(\alpha) & t_x \\ \sin(\alpha) & \cos(\alpha) + s & t_y \\ 0 & 0 & 1 \end{pmatrix}, \\ \alpha &= \frac{\Pi}{16} \cdot r_k \cdot w \qquad \text{(Rotationsfehler)} \\ s &= 0.2 \cdot r_k \cdot w \qquad \text{(Skalierungsfehler)} \\ t_x &= 9 \cdot r_k \cdot w \qquad \text{(Translationsfehler)} \\ t_y &= 9 \cdot r_k \cdot w \end{aligned} \tag{10}$$

und der standardnormalverteilten Zufallszahl r_k wird in Abbildung 3 und 4 verdeutlicht, dass die visuelle Wahrnehmung der Registrierung durch den $mean_{C_L}$ gut wiedergegeben wird.

Das Verfahren von Kourogi kann durch die Optimierungen entscheident verbessert werden und liefert relativ gute Ergebnisse. Der neue Algorithmus erreicht bessere Ergebnisse, ist stabiler und die Laufzeit schneller als die des zuvor optimierten Verfahrens von Kourogi (vgl. Tabelle 1 und 2).

In Abbildung 5, 6 und 7 ist das stabilisierte Resultat von zwei Sequenzen durch den neuen Algorithmus dargestellt. Die Abbildung 6 verdeutlicht, dass das Aliasing-Artefakt im Testmuster durch die Stabilisierung kompensiert wird und Abbildung 7, dass die Auflösung des Bildes erhöht wurde. Tabelle 3 zeigt die Verbesserung der Wiedererkennung eines stabilisierten Bildes.

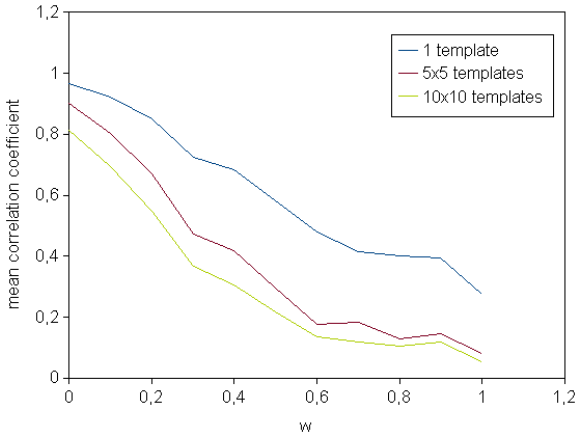


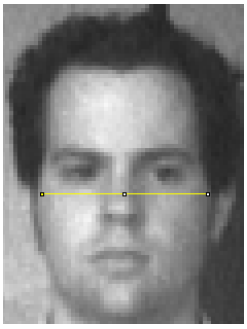
Abbildung 4: Durchschnittlicher $mean_{C_L}$ für das gesamte Testset in Abhängigkeit von dem künstlichen Fehlergewicht w. Ein Template bildet visuell schlechte Matches noch mit einem $mean_{C_L} = 0.3$ ab. 10×10 Templates bewerten visuell gute Matches mit durchschnittlich $mean_{C_L} = 0,8$. 5×5 Templates geben den visuellen Eindruck am besten wieder.

Method	$mean_{C_L}$	
	mean	std
Kourogi in Bounding Box	0,53	0,38
Kourogi in Boundig Box+0.25%, IIR $a = 4.5$, $T_{maxmotion} = 15$	0,84	0,18
Log-Search in Bounding Box with initial transformation	0,89	0,09

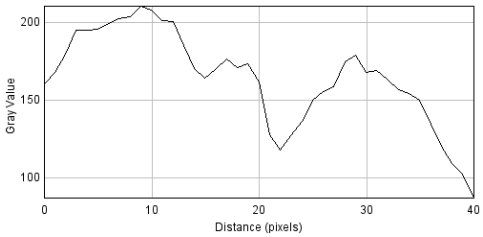
Tabelle 1: Mittelwert und Standardabweichung des $mean_{C_L}$ für das Testset.

Method	**mean**	**std**
Kourogi in Boundig Box+0.25%, IIR $a = 4.5$, $T_{maxmotion} = 15$	8,34	7,22
Log-Search in Bounding Box with initial transformation	4,84	2,12

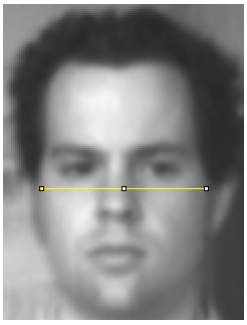
Tabelle 2: Laufzeitvergleich: Durchschnitt und Standardabweichung der Laufzeit der Matlab-Implementierung in Sekunden (System: Intel Core 2 Duo 1,86 GHz E6320, 1 GB RAM).



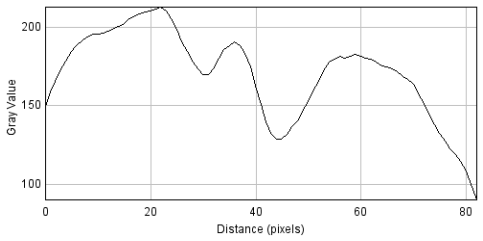
(a) Referenzbild



(b) Profilplot des Referenzbildes



(c) 2×skaliertes Average



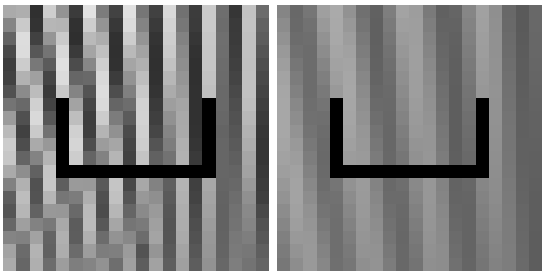
(d) Profilplot des Average Bildes

Abbildung 5: Vergleich 2× Average-Verfahren und Referenzbild der Sequenz *n009a* mit 10 Bildern (Registrierung wurde mit dem neuen Algorithmus durchgeführt)

(a) Referenzbild (2× vergrößert ohne Interpolation)

(b) Average 2×

Abbildung 6: Averaging mit Testmuster (Sequenz *sr016b* mit 21 Bildern): Die Bilder zeigen das Referenzbild der Sequenz in Gegenüberstellung zu dem Average Bild. Das Testmuster verdeutlicht, dass die 2× skalierten Average Bilder eine höhere Auflösung aufweisen, als das Referenzbild.



(a) Referenzbild (b) 2× Average

Abbildung 7: Pixelvergleich Referenzbild - 2× Average Bild: Links ist das Referenzbild und rechts das 2× Average Bild der Sequenz *sr016b* zu sehen. In den Bildern ist jeweils ein Ausschnitt von 10 Pixeln der gleichen Position im Testmuster abgebildet. Das Average Bild bildet mit 10 Pixeln zwei schwarze Linien und das Referenzbild circa 4 schwarze Linien des Testmusters ab.

Image	**Similarity**
1	0,382
2	0,488
3	0,436
4	0,436
5	0,424
6	0,385
7	0,348
Average	0,503
2×Average	0,507

Tabelle 3: Verbesserung der Gesichtswiedererkennung durch die Average Bilder: Die Tabelle zeigt die erziehlte Ähnlichkeit der Einzelbilder der Sequenz im Comparision Window des FaceCheckers (CrossMatch Technologies Inc.) und der Average Bilder mit dem Potraitbild eines Probanden. Das Average Bild und 2× skalierte Average Bild wird den Einzelbildern der Sequenz gegenübergestellt. Die Average Bilder verbessern die Wiedererkennung.

5 Fazit

Die Untersuchung der Verfahren anhand der Testaufnahmen hat gezeigt, dass die intensitätsbasierte Registrierung [3] auf Grundlage von einzelnen Pixeln in einer Region, die eine grobe Schätzung für den registrierungsrelevanten Bereich darstellt, durch die Optimierungen verbessert werden kann. Das Verfahren ist jedoch nicht robust genug gegenüber Rauschen und Intensitätsdifferenzen. Der neue Algorithmus [5] erreicht durch die Suche mittels normalisierter Kreuzkorrelation und Logarithmic Search gute Resultate.

Für den Vergleich der registrierten Bilder wurden verschiedene Vergleichsmaße entwickelt, deren Parameter durch systematische Variation bestmöglich ermittelt und auf umfangreichem Testmaterial untereinander und gemäß ihres visuellen Eindrucks verglichen wurden. Ein neues Vergleichsmaß auf Basis des normierten Korrelationskoeffizenten, gibt den visuellen Eindruck der Registrierung gut wieder.

Der neue Algorithmus [5] ist geignet um Superresolution-Bilder zu erzeugen und die Laufzeit der Matlab-Implementierung ist im Durchschnitt besser als die des zuvor optimierten Verfahrens von Kourogi [3].

Die Erkennungsleistung wird durch das stabilisierte Average Bild deutlich verbessert.

References

[1] David Capel: Image Mosaicing and Superresolution, Springer-Verlag, London, 2004.

[2] Jürgen Fuchs: Warum die Bildqualität der Uberwachungsanlagen den Ermittler nicht immer begeister, Polizei heute 3/2007, Richard Boorberg Verlag, S. 119-121, 2007.

[3] Masakatsu Kourogi et. al.: Real-Time Image Mosaicing From a Video Sequence., In: Procs ICIP99, vol.4, S. 133-137, 1999.

[4] Wolfgang Konen et al.: Endoscopic image mosaices for real-time color video sequences, In: H.U. Lemke (ed.), Computer Assisted Radiology and Surgery (CARS2007, Berlin), Elsevier, Amsterdam, 2007.

[5] Wolfgang Konen: persönliche Mitteilung, Dezember 2008.

[6] Scott Shald: Comparision of FIR and IIR Filters in Coherent Lidar Processing, 14th Coherent Laser Radar Conference, Snowmas, Colorado, USA, August 2007

[7] J.R. Jain and A.K. Jain: Displacement measurement and its application in interframe image coding, IEEE Transactions on Communications, vol. COM-29, S. 1799-1808, Dec 1981.