Morphable Face Models – Erzeugung und Anwendungen

David C. Schneider, Peter Eisert

Fraunhofer Institut für Nachrichtentechnik HHI
Einsteinufer 37
10587 Berlin
david.schneider@hhi.fraunhofer.de
peter.eisert@hhi.fraunhofer.de

Zusammenfassung Wir beschreiben ein parametrisches dreidimensionales Modell des menschlichen Gesichtes, *Morphable Face Model* genannt, das mit statistischen Methoden aus einer Datenbank gescannter Gesichter generiert wird. Wir zeigen, dass zur Erzeugung dieses Modells eine semantisch konsistente Registrierung von Gesichtsmodellen notwendig ist und stellen einen vollautomatischen Algorithmus hierfür vor.

1 Einleitung

Ein *Morphable Face Model*, im Folgenden kurz MFM, ist ein generisches dreidimensionales Modell des menschlichen Gesichts. Ein MFM wird durch eine Parametermatrix beschrieben deren Einträge die Ausprägung verschiedener Gesichtsmerkmale festlegen, etwa die Form der Nase oder die Breite des Kinns. Für das Gesicht einer realen Person kann ein Parametersatz berechnet werden, der das MFM an die Erscheinung dieser Person anpasst. Die Parameter des MFM werden nicht manuell festgelegt, sondern vielmehr mit statistischen Methoden aus einer Datenbank von Modellen echter Gesichter gelernt. Dadurch ist sichergestellt, dass die Parameter tatsächlich auftretende Variationen von Gesichtern beschreiben.

MFMs sind eine Repräsentation menschlicher Gesichter mit vielfältigen Vorzügen, von denen hier nur einige aufgeführt werden sollen. So stellen sie eine besonders kompakte Beschreibung von Gesichtsformen dar. Anstelle eines Gittermodells mit tausenden von Dreiecken wird das Gesicht durch die Parametermatrix des Modells beschrieben, die zum Beispiel im Falle des von den Autoren entwickelten MFM lediglich 25 Einträge umfasst.

Ferner sind MFMs – im Gegensatz zu den Punktmodellen, die etwa ein Laserscanner erzeugt – semantisch annotiert. Das bedeutet, dass topologische äquivalente Teile in MFM-Darstellungen verschiedener Personen denselben Teil des jeweiligen Gesichts repräsentieren, z.B. die Nasenspitze oder die Mitte der

Oberlippe, auch wenn diese Teile von Person zu Person unterschiedlich geformt sind.

MFMs haben eine Vielzahl nützlicher Anwendungen in der Computer Vision und Computergrafik. Einige Beispiele:

- In Schneider und Eisert (2008) zeigen wir, wie mit Hilfe eines MFM Fehler in dreidimensionalen Gesichtsscans repariert werden können. Auch wenn lediglich 20% der Messdaten fehlerfrei sind, kann mit einem MFM eine plausible Rekonstruktion des Gesichts berechnet werden.
- Paterson und Fitzgibbon (2003) verwenden ein MFM für das Echtzeit-Tracking von Köpfen in drei Dimensionen.
- Vetter und Blanz (1998), schätzen mit einem MFM dreidimensionale Gesichtsmodelle von Fotos.
- Wird eine Animation für die Topologie eines MFM definiert, so kann jedes Gesicht, sobald es durch das MFM beschrieben wurde, dementsprechend animiert werden, trotz der unterschiedlichen Form verschiedener Gesichter.

Im Folgenden beschreiben wir eine neue Methode zur vollautomatischen Erstellung von MFMs aus einer Datenbank dreidimensionaler Gesichtsmodelle, die durch einen Laserscanner oder eine vergleichbare Technologie zur dreidimensionalen Formerfassung (z.B. Fechteler et al., 2007) erzeugt wurden. Dabei erfolgt die statistische Berechnung des MFM weitgehend nach der Standardmethodik von Blanz und Vetter (1999). Vor der eigentlichen Berechnung müssen die Modelle aus der Datenbank jedoch semantisch konsistent registriert werden. Dazu beschreiben wir in diesem Artikel ein neues Verfahren.

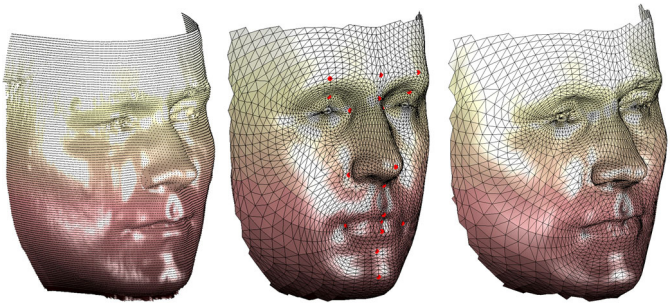

Abbildung 1 Links: Laserscan. Mitte: Referenzmodell mit Landmarken. Rechts: Laserscan nach semantischer Registrierung mit Referenzmodell.

2 Semantische Registrierung

Was bedeutet „semantisch konsistente Registrierung“? Ein Gesichtsmodell, das mit einem Laserscanner erstellt wird, ist in der Regel eine dichte Punktwolke, siehe Abb. 1 links. Modelle verschiedener Personen variieren in der Anzahl der Punkte und topologisch äquivalente Punkte in verschiedenen Modellen sind semantisch nicht aufeinander bezogen, d.h. sie beschreiben verschiedene Teile der jeweiligen Gesichter. Zur Erstellung eines MFM müssen nun alle Modelle der Datenbank so transformiert werden, dass sie dieselbe Topologie aufweisen und dass topologisch äquivalente Punkte dieselben Teile der jeweiligen Gesichter beschreiben. Ohne diese semantische Konsistenz würden verschiedene Teile des Gesichts bei der Erzeugung des MFM statistisch vermengt und es könnte kein aussagekräftiges Modell erstellt werden.

Für die automatische semantische Registrierung wird ein Referenzmodell verwendet, das die Topologie des MFM festlegt. Als Referenzmodell wird ein „neutrales“ Gesicht ohne besondere charakteristische Merkmale verwendet; siehe Abb. 1 Mitte. Dieselbe Abbildung zeigt rechts das Ergebnis der semantischen

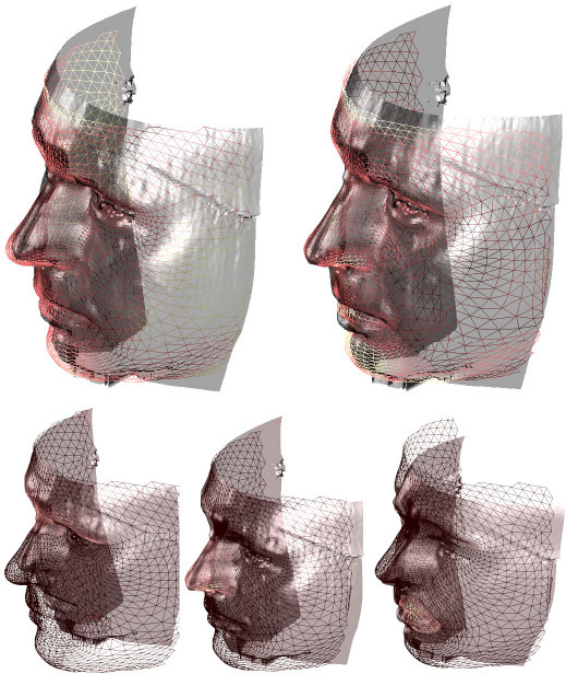

Abbildung 2 Stufen der ICP-Registrierung.
Oben: Starre und affine Transformation.
Unten: Lokale affine Transformationen für drei verschiedene Landmarken.

Registrierung des links dargestellten Laserscans. Die automatische Registrierung erfolgt in drei Schritten.

Im ersten Schritt werden mit Hilfe des Referenzmodells n charakteristische Punkte sogenannte Landmarken, im Laserscan aus der Datenbank ermittelt; siehe Abb. 1 Mitte. Die Landmarken sind im Referenzmodell als Punkte $\mathbf{p}_1 \ldots \mathbf{p}_n$ annotiert. Dazu wird eine mehrstufige Variante des *Iterative Closest Points*-Algorithmus (ICP) eingesetzt.

ICP berechnet iterativ eine Transformation **T**, die eine dreidimensionale Punktwolke **A** möglichst genau an einer anderen Punktwolke **B** ausrichtet (s. z.B. Rusinkiewicz und Levoy, 2001). Die Schwierigkeit besteht darin, dass die ideale Transformation davon abhängt, welche Punkte in **A** an welchen Punkten in **B** auszurichten sind. Daher wird angenommen, dass jeder Punkt in **A** am jeweils euklidisch nächsten Punkt in **B** auszurichten ist. Aus dieser Korrespondenz kann ein optimales **T** berechnet und auf **A** angewendet werden. **A** rückt damit näher an **B** heran und die Prozedur wird wiederholt. Es ist bewiesen, dass der Algorithmus konvergiert (Besl und McKay, 1992), allerdings nicht unbedingt zur idealen Transformation **T**. Das Konvergenzverhalten hängt zum einen davon ab, wie weit die Punktwolken von der idealen Ausrichtung entfernt sind, zum anderen davon, wie viele freie Parameter die verwendete Transformation **T** hat. Daher wird ICP üblicherweise mit einer starren Bewegung verwendet.

Zur Anpassung der Landmarken verwenden wir ein dreistufiges modifiziertes ICP-Verfahren. Dazu wird zunächst eine starre Bewegung $\mathbf{T}_\mathbf{r}$ des Referenzmodells zum Laserscan ermittelt; siehe Abb. 2 oben links. Dadurch werden Scan und Referenzmodell hinreichend präzise ausgerichtet um anschließend eine affine Transformation $\mathbf{T}_\mathbf{a}$ zu berechnen; siehe Abb. 2 oben rechts. Diese Transformation hat eine höhere Zahl von Freiheitsgraden und wäre ohne vorhergehende starre Ausrichtung mit ICP nicht stabil zu ermitteln. Anschließend wird das Referenzmodell in n Teilmodelle $\mathbf{R}_1 \ldots \mathbf{R}_n$ zerlegt; ein Teilmodell $\mathbf{R}_i$ ist dabei die lokale Umgebung der Landmarke $\mathbf{p}_i$. Jedes Teilmodelle $\mathbf{R}_i$ wird schließlich mit einem affinen ICP-Durchlauf weiter an den Laserscan angepasst; daraus ergeben sich n weitere affine Transformationen $\mathbf{T}_{\mathbf{a},1} \ldots \mathbf{T}_{\mathbf{a},n}$ (siehe Abb. 2 unten). Damit ist die Lage $\mathbf{q}_i$ der Landmarke i im Laserscan bestimmt als

$$\mathbf{q}_i = \mathbf{T}_{\mathbf{a},i} \mathbf{T}_\mathbf{a} \mathbf{T}_\mathbf{r} \mathbf{p}_i \,.$$

Abb. 3 zeigt einige Beispiele automatisch platzierter Landmarken.

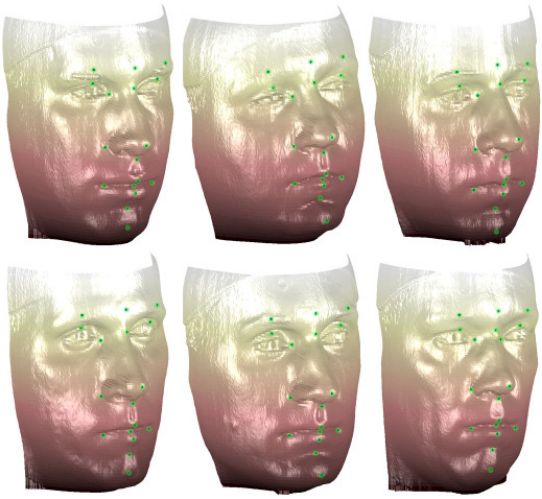

Abbildung 3 Beispiele für automatisch platzierte Landmarken in sechs Laserscans.

Die Transformationen aus dem oben geschilderten Schritt sind lokal. Jede der n Transformationen bildet genau eine Landmarke des Referenzmodells auf die entsprechende Stelle im Scan ab, sie erlaubt aber keine Aussage über Korrespondenzen von Punkten zwischen den Landmarken. Daher wird im zweiten Schritt mit Hilfe der Landmarken eine einzige globale, nichtlineare Verzerrung berechnet, welche die Landmarken im Laserscan auf die Landmarken im Referenzmodell abbildet und für alle Punkte, die keine Landmarken sind, eine passende Abbildung interpoliert. Diese soll die ursprüngliche Geometrie möglichst wenig verzerren, gleichzeitig aber keine künstlichen Diskontinuitäten herbeiführen. Diese Anforderungen erfüllt der aus der zweidimensionalen Bildregistrierung bekannte Formalismus der *Thin Plate Splines* (TPS), der aus der Minimierung einer physikalisch interpretierbaren Biegeenergie hergeleitet werden kann (Bookstein, 1989).

Im Gegensatz zur Bildregistrierung haben wir hier ein dreidimensionales Interpolationsproblem mit unregelmäßig verteilten Stützpunkten zu lösen. Die bekannten Größen sind die Verschiebungsvektoren

$$\mathbf{d}_i = \mathbf{p}_i - \mathbf{q}_i ,$$

die eine Landmarke im Laserscan auf eine Landmarke im Referenzmodell abbilden. Zu interpolieren sind die Verschiebungsvektoren derjenigen Punkte, die keine Landmarken sind. Für ein dreidimensionales Problem ist die radiale Basisfunktion der TPS die Betragsfunktion. Mit

$$u_{i,j} = \left\| \mathbf{p}_i - \mathbf{p}_j \right\|$$

erhalten wir eine TPS-Matrix

$$\mathbf{S} = \begin{bmatrix} 0 & u_{1,2} & \cdots & u_{1,k} & 1 & \mathbf{q}_1^T \\ u_{2,1} & 0 & & u_{2,k} & 1 & \mathbf{q}_2^T \\ \vdots & & \ddots & \vdots & 1 & \vdots \\ u_{k,1} & u_{k,2} & \cdots & 0 & 1 & \mathbf{q}_k^T \\ 1 & 1 & \cdots & 1 & 0 & \\ \mathbf{q}_1 & \mathbf{q}_2 & \cdots & \mathbf{q}_k & & \mathbf{0}_{3\times 3} \end{bmatrix}$$

Damit erhalten wir als Gewichtsmatrix für die TPS-Transformation

$$\mathbf{W} = \mathbf{S}^{-1} \left[\mathbf{d}_1 \cdots \mathbf{d}_n \; \mathbf{0}_{3\times 3} \right]^T$$

und ein beliebiger Punkt **s** wird abgebildet auf

$$\mathbf{s}' = \mathbf{s} + \mathbf{W}^T \left[u_1 \cdots u_k \; 1 \; \mathbf{p}^T \right]^T$$

wobei

$$u_i = \left\| \mathbf{s} - \mathbf{q}_j \right\|.$$

Nach der Landmarken-geleiteten TPS-Deformation liegen semantisch äquivalente Bereiche des (deformierten) Laserscans und des Referenzmodells nun übereinander. Der Laserscan ist aber immer noch eine dichte Punktwolke. Im letzten Schritt der Registrierung wird der Scan daher in der Topologie des Referenzmodells neu aufgelöst. Dazu wird der Laserscan zunächst als Dreiecksnetz repräsentiert und zur Beschleunigung der folgenden Berechnungen in einem AABB-Baum (*axis aligned bounding boxes*) abgespeichert. Aufgrund der hohen Anzahl von Dreiecken, die sich aus den ca. 35000 Punkten eines Scans ergibt, wird der Baum nur bis zu einer Tiefe von 200 Dreiecken pro Blatt ausgeführt.

Anschließend wird für jeden Punkt **p** des Referenzmodells eine Gerade durch **p** in Richtung der Oberflächennormalen mit dem Laserscan geschnitten. Der Schnittpunkt mit dem geringsten Abstand zu **p** wird als semantisches Äquivalent zu **p** im *deformierten* Laserscan betrachtet. Letztendlich sollen aber die Korrespondenzen zwischen dem Referenzmodell und dem ursprünglichen, nicht deformierten Laserscan berechnet werden. Daher wird jeder Schnittpunkt über den Index des geschnittenen Dreiecks und seine baryzentrischen Koordinaten darin beschrieben. Damit kann seine Lage im ursprünglichen Laserscan, der nun ebenfalls als Dreiecksnetz betrachtet wird, rekonstruiert werden. Das Endergebnis der Registrierung und Topologieübertragung zeigt Abbildung 4.

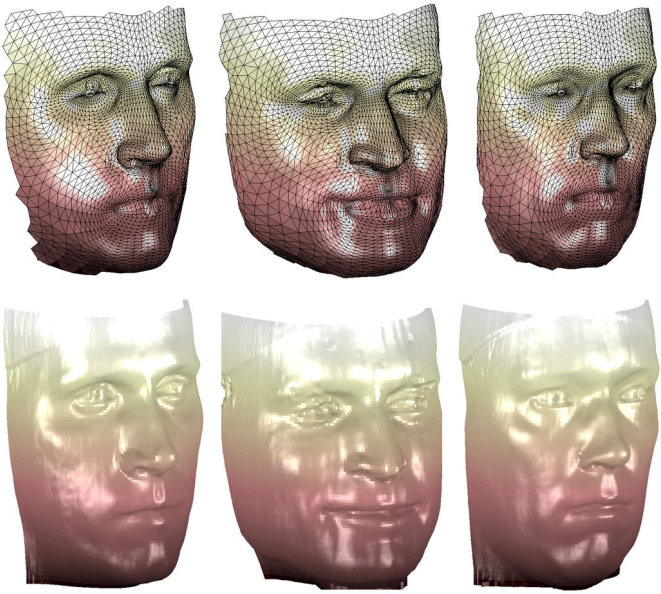

Abbildung 4 Ergebnisse der semantischen Registrierung

3 Modellgenerierung

Aus einer Datenbank semantisch registrierter Kopfmodelle kann nun das MFM generiert werden. Dazu wird das Gesicht in fünf Regionen unterteilt, die sich an den Rändern überlappen (Augenpartie, Nase, Mund, Kinn, Rest). Jede einzelne Region wird durch ein lineares Modell beschrieben, das durch Hauptkomponentenanalyse (PCA) gewonnen wird. Dabei wird jeder Punkt im Gittermodell als eine Variable behandelt. Das Modell wird durch je eine Hauptkomponentenmatrix $\mathbf{P}_k$ und einen Mittelwertsvektor $\boldsymbol{\varphi}_k$ pro Gesichtsregion repräsentiert.

Um ein gegebenes Gesicht durch das MFM zu beschreiben, wird es mit dem oben beschriebenen Verfahren registriert und wie folgt in den Modellraum projiziert: Es sei $\mathbf{m}_k$ der Vektor, die Geometrie der *k*ten Gesichtsregion beschreibt. Die Parametermatrix $\mathbf{W}$ ist dann gegeben durch

$$\mathbf{W} = [\mathbf{P}_1^+(\mathbf{m}_1 - \boldsymbol{\varphi}_1) \ldots \mathbf{P}_k^+(\mathbf{m}_k - \boldsymbol{\varphi}_k)]$$

Dabei bezeichnet $\mathbf{P}^+$ die Pseudoinverse einer Matrix $\mathbf{P}$. Umgekehrt wird aus einer Parametermatrix $\mathbf{W}$ mit Spaltenvektoren $\mathbf{w}_k$ die Geometrie $\mathbf{g}_k$ der *k*ten Gesichtsregion rekonstruiert durch

$$\mathbf{g}_k = \boldsymbol{\varphi}_k + \mathbf{P}_k \mathbf{w}_k .$$

Die Genauigkeit, mit der das Modell an gegebene Gesichter angepasst werden kann, wird durch die Anzahl der verwendeten Hauptkomponenten, d.h. durch die Spaltenzahl der Matrizen $\mathbf{P}_k$, bestimmt. Die Größe der Parametermatrix ist gegeben durch die Zahl der verwendeten Hauptkomponenten mal der Anzahl der Gesichtsregionen. Fünf Hauptkomponenten pro Region und damit eine Parametermatrix mit 25 Elementen ist ausreichend um Gesichter durch das MFM ohne erkennbaren Unterschied zum Gittermodell zu beschreiben.

4 Zusammenfassung und Ausblick

Wir haben *Morphable Face Models*, ein parametrisches dreidimensionales Modell des menschlichen Gesichtes, beschrieben, das mit statistischen Methoden aus einer Datenbank gescannter Gesichter generiert wird. Wir haben gezeigt, dass zur Erzeugung dieses Modells eine semantisch konsistente Registrierung von Gesichtsmodellen notwendig ist und einen robusten, vollautomatischen Algorithmus hierfür vorgestellt. Dieser beruht auf einer automatischen Bestimmung von Landmarken im gescannten Gesicht, einer nichtlinearen Abbildung des Scans auf ein Referenzmodell und der Übertragung des Scans in dessen Topologie.

Nicht behandelt haben wir hier die Textur des Gesichts, die z.B. bei Blanz und Vetter (1999) in das Modell mit einfließt. Tatsächlich wurde die die Hauptkomponentenanalyse, die statistische Methode, die den MFMs zugrunde liegt, schon lange vorher erfolgreich auf Bilder von Gesichtern angewandt, siehe Turk und Pentland (1991). Die Erweiterung unseres Modells um Texturdaten ist geplant.

MFMs können zur Verbesserung der Ergebnisse dreidimensionaler Gesichtserfassung eingesetzt werden. Damit können auch mit Methoden, die weniger kostenintensiv sind als Laserscanner, Gesichter robust dreidimensional erfasst werden; siehe z.B. Fechteler et al. (2007), wo eine handelsübliche Hardware zur 3D-Erfassung eingesetzt wird. In Schneider und Eisert (2008) beschreiben wir allgemein die Anwendung des MFM zur Reparatur defekter Laserscans. Diese Methoden können ausgebaut und in den Prozess der dreidimensionalen Datenerfassung integriert werden.

Referenzen

Paul J. Besl, Neil D. McKay. *A method for registration of 3-d shapes.* IEEE Transactions on Pattern Analysis and Machine Intelligence, 14:239–256, 1992.

Volker Blanz, Thomas Vetter. *A morphable model for the synthesis of 3d faces.* Proceedings of the 26th annual conference on Computer graphics and interactive techniques, 1999.

F. L. Bookstein. *Principal warps: Thin-plate splines and the decomposition of deformations.* IEEE Transactions on Pattern Analysis and Machine Intelligence, 11:567 – 585, 1989

Philipp Fechteler, Peter Eisert, Jürgen Rurainsky. *Fast and high resolution 3D face scanning.* Proc. International Conference on Image Processing, San Antonio, USA, 2007.

J. Paterson, A. Fitzgibbon. *3D head tracking using non-linear optimization.* British Machine Vision Conference 03, 2003.

D. C. Schneider, P. Eisert: *Automatic and Robust Semantic Registration of 3D Head Scans*, 5th European Conference on Visual Media Production (CVMP), London, 2008

Szymon Rusinkiewicz, Marc Levoy. *Efficient variants of the ICP algorithm.* Third International Conference on 3D Digital Imaging and Modeling, 2001.

Matthew Turk and Alex Pentland. *Face recognition using Eigenfaces*. Proc. IEEE Conference on Computer Vision and Pattern Recognition, 1991.

Thomas Vetter, Volker Blanz. *Estimating coloured 3D face models from single images: An example based approach.* Computer Vision - ECCV 98, Volume 1407/1998 of Lecture Notes in Computer Science. Springer, 1998.