

# 3D-Rekonstruktion von Objekten mittels „Structure-from-Motion“ aus einer photogrammetrischen Aufnahme mit den Programmen VisualSFM und CPMVS

MARIUS ZIEGLER<sup>1</sup>, EBERHARD GÜLCH<sup>2</sup> & PAUL RAWIEL<sup>3</sup>

*Zusammenfassung: Es gibt Programme, mit denen aus Mehrbildaufnahmen eines Objektes eine 3D-Punktwolke berechnet und daraus ein 3D-Modell erstellt werden kann. Die Programme werden allgemein als „Structure-from-Motion“-Programme bezeichnet. Im Mittelpunkt steht das Programm VisualSFM. VisualSFM detektiert Merkmale in den Bildern, verknüpft diese und rekonstruiert daraus das Objekt als dünne Punktwolke. Für die Verdichtung der Punktwolke sind zwei Stufen vorgesehen. Mit den Programmen CMVS/PMVS2 erhält man in kurzer Zeit ein einfaches 3D-Modell. Zeitaufwändiger ist die Verwendung von CPMVS, dafür ist das Resultat ein hochwertiges und detailliertes Modell. Diese Arbeit untersucht die Funktionalität dieser Prozessschritte und vergleicht das Ergebnis mit dem Alternativprogramm 123D-Catch.*

*VisualSFM ermöglicht eine Einflussnahme auf die Berechnungen. Mit Funktionen wie z.B. dem „Sequence Matching“ kann die Auswertung nicht nur beschleunigt sondern auch Fehler vermieden werden. Die Korrektur der Objektivverzeichnung wird von VisualSFM allein anhand der importierten Fotos durchgeführt. Besonders wichtig sind die Exif-Informationen der Fotos, die u.a. die Objektivbrennweite der Kamera enthalten. Fehlt diese, berechnet VisualSFM nach einer festgelegten Formel eine Brennweite, die meist viel zu groß ausfällt. Dies führt zu mehr oder weniger starken Deformationen. Mit 123D-Catch ließen sich dabei keine brauchbaren 3D-Modelle mehr berechnen.*

## 1 Einleitung

Es ist heutzutage relativ einfach, ein reales Objekt als 3D-Modell am PC abzubilden. Mit der Rechenleistung moderner Computer können selbst größere Datenmengen in Minuten oder wenigen Stunden verarbeitet werden. Zu den üblichen Verfahren zur Datengewinnung zählen das Laserscanning und die Photogrammetrie. Laserscanner sind teuer, benötigen einiges an Übung und sind daher nur im professionellen, gewerblichen Umfeld im Einsatz. Im Gegensatz dazu befinden sich digitale Kameras in jedem Mobiltelefon. Der Vorteil der Photogrammetrie liegt in der einfachen Anwendung, die jeder Person in kurzer Zeit erklärt werden kann. Ist das Objekt aus verschiedenen Perspektiven fotografiert, lässt sich dieses rechnerisch rekonstruieren. Dies führt zu der Bezeichnung „Structure-from-Motion“ (kurz „SfM“). Die Rekonstruktion der „Structure“ durch die „Motion“, also die Bewegung der Kamera zwischen den verschiedenen Aufnahmestandpunkten. Die Rekonstruktion erfolgt mit Programmen am Computer. Im

---

<sup>1</sup> M.Eng. Marius Ziegler, HFT Stuttgart, Schellingstr. 24, 70174 Stuttgart; E-Mail: [marius.ziegler@hft-stuttgart.de](mailto:marius.ziegler@hft-stuttgart.de)

<sup>2</sup> Prof. Dr. Eberhard Gülch, HFT Stuttgart, Schellingstr. 24, 70174 Stuttgart; E-Mail: [eberhard.guelch@hft-stuttgart.de](mailto:eberhard.guelch@hft-stuttgart.de)

<sup>3</sup> Prof. Dr. Paul Rawiel, HFT Stuttgart, Schellingstr. 24, 70174 Stuttgart; E-Mail: [paul.rawiel@hft-stuttgart.de](mailto:paul.rawiel@hft-stuttgart.de)

Mittelpunkt der Untersuchungen steht das Programm VisualSFM (Wu C. , 2012). Zum Vergleich wird das Programm 123D-Catch (Autodesk, 2012) herangezogen.

Das Ziel dieser Arbeit ist die Untersuchung der Leistungsfähigkeit von VisualSFM und des Vergleichsprogramms 123D-Catch. Dazu werden die verschiedenen Schritte des Auswertungsprozess näher betrachtet. Wichtige Eigenschaften und Möglichkeiten der Einflussnahme werden beschrieben, so dass zukünftige Untersuchungen auf diesem Wissen aufbauen können.

Die empirischen Untersuchungen wurden mit zwei Kameras, einer Canon 400D DSLR und einer Canon G12 Kompaktkamera, durchgeführt. Datensätze wurden von einem kleinen Schloss, einem zwei Meter Würfel und einem kleinen Stück Fassade angefertigt.

## 2 VisualSFM

Das Programm wird über eine grafische Benutzeroberfläche gestartet. Im Standardfall sind drei Aktionen nacheinander auszuführen, bis man als Ergebnis eine dünne Punktwolke erhält. Im vierten Schritt wird die Punktwolke verdichtet und ein geschlossenes Modell berechnet.

### 2.1 Abmessung der Fotos

Je größer die Abmessung (in Pixel) der Fotos, desto mehr Arbeitsspeicher beansprucht VisualSFM. Reicht dieser nicht aus, wird die Anwendung instabil und stürzt ab. Damit genügend Ressourcen zur Verfügung stehen, sollte ein PC mit 64Bit-Betriebssystem und großzügigem Arbeitsspeicher verwendet werden. Eine Reduzierung der Auflösung der Fotos ist nur bedingt zu empfehlen. Zwar wird dadurch die Berechnung insgesamt beschleunigt, allerdings auch die Wahrscheinlichkeit erhöht, ein deformiertes bzw. fehlerhaftes 3D-Modell zu erhalten.

Für die Merkmalsextraktion wird der Sift-Algorithmus (Lowe, 1999) verwendet. Standardmäßig wird hier „SiftGPU“ (Wu C. , 2007), einer Implementierung des Sift-Algorithmus unter Verwendung der GPU, dem Prozessor der Grafikkarte, verwendet. Dabei ist die größere Seitenlänge, also die Höhe oder Breite des Fotos, programmseitig auf 3200 Pixel begrenzt. Zur Vorbereitung für „SiftGPU“ wird die Auflösung der Fotos grundsätzlich zuerst verdoppelt. Anschließend werden die Fotos unter Wahrung des Seitenverhältnisses nach einer festen Formel so angepasst, dass die größere Seitenlänge möglichst nah an, aber unter 3200 Pixel liegt. Das bedeutet, dass große Bilder ggf. verkleinert werden und kleine Bilder vergrößert. VisualSFM bietet alternativ auch eine Merkmalsdetektion mit Sift auf der CPU durchzuführen, allerdings konnte diese Funktion nicht erfolgreich verwendet werden.

### 2.2 Matchingverfahren

Im Normalfall führt VisualSFM das Matching automatisch zwischen allen importierten Bildern durch. Bei kleinen Datensätzen mit eindeutigen Bildmerkmalen ist diese Vorgehensweise nicht zu beanstanden. Problematisch wird dies aber zum einen bei großen Datensätzen, da in diesem Fall der Berechnungsaufwand mit jedem Bild überproportional zunimmt, zum anderen wenn Fotos unterschiedliche Bereiche eines Objektes abbilden und auf Grund ähnlicher Strukturen eine fehlerhafte Zuordnung entsteht (vgl. Abb. 1 links).

Ist die Zuordnung fehlerhaft, bietet VisualSFM die Möglichkeit, mit „Resume 3D-Reconstruction“ das Matching wiederaufzunehmen. Das Programm versucht weitere

Verknüpfungen zwischen den Fotos zu bilden und bestehende zu verbessern. Führt dies auch nach mehrfacher Wiederaufnahme nicht zum Erfolg, muss das Matching komplett neu gestartet werden. Dabei besteht die Möglichkeit, das „Initial Pair“, also das Bildpaar, mit dem VisualSFM das Matching beginnt und dem alle anderen angepasst werden, manuell festzulegen.

Bei bestimmten Aufnahmeconfigurationen ist das „Sequenz Matching“ eine Alternative. Das Prinzip dahinter ist eine Eingrenzung des Matching auf benachbarte Bilder. Dadurch kann ein fehlerhaftes Matching komplett verschiedener Objektansichten vermieden (vgl. Abb. 1 rechts) und das Matching bei großen Datensätzen erheblich beschleunigt werden. Rechenbeispiel: Bei einem Datensatz von 76 Fotos sind bei einem kompletten Matching 2880 Vergleiche erforderlich, beim „Sequenz Matching“ mit jeweils 4 benachbarten Bildern nur 152.

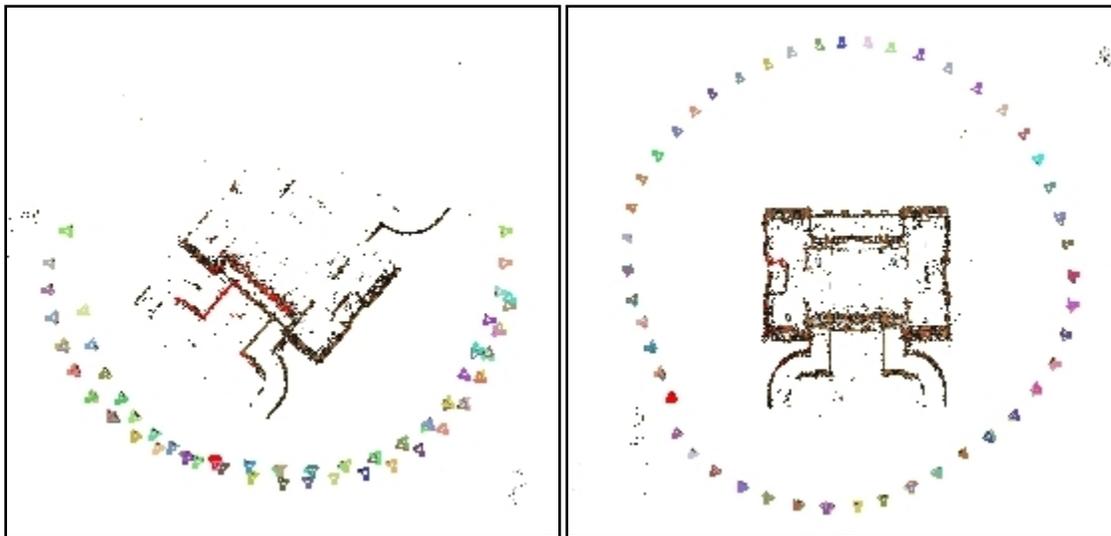


Abb. 1: Zwei Ergebnisse der 3D-Rekonstruktionen mit VisualSFM. Links fehlerhaftes Modell nach „komplettem“ Matching, rechts korrektes Ergebnis mit „Sequenz Matching“.

VisualSFM verwendet für das Matching die GPU der Grafikkarte. Dafür stehen zwei Technologien zur Auswahl. Standardmäßig wird GLSL (OpenGL Shading Language) verwendet, da diese plattformunabhängig ist. Speziell für NVIDIA Grafikkarten wurde von NVIDIA die CUDA Technik entwickelt (NVIDIA GmbH, 2012). Beide Techniken verfolgen das gleiche Prinzip, für besonders aufwendige Berechnungen nicht allein die CPU, sondern zusätzlich die GPU der Grafikkarte zu nutzen. Bei einer Laufzeitmessung mit 70 Fotos vom importieren bis zur berechneten dünnen Punktwolke ergab keine nennenswerten Unterschiede. Beide Technologien können in diesem Anwendungsfall als gleichwertig angesehen werden.

### 2.3 Korrektur der Kameraverzeichnung

Im Regelfall sind die in VisualSFM importierten Fotos nicht frei von Objektivverzeichnung. Während des Auswerteprozesses entzerrt VisualSFM die Fotos allein anhand der Bildinformationen. Ein Vergleich der Fotos vor und nach der Entzerrung hat gezeigt, dass dabei lediglich die radialsymmetrische Verzeichnung des Objektivs korrigiert wird. Das Hinzufügen eigener Kalibrierungsdaten ist über eine Eingabemaske über das Menü möglich. Es hat sich aber herausgestellt, dass dann von VisualSFM keine Entzerrung der Fotos mehr durchgeführt wird. Es

ist in diesem Fall also erforderlich, dass der Anwender die Fotos vor der Auswertung mit VisualSFM selbst entzerzt, was den Aufwand der Auswertung deutlich erhöht.

## 2.4 Exif-Informationen

Bei der Aufnahme von Fotos mit einer handelsüblichen Digitalkamera speichert diese verschiedene Metadaten der Aufnahmesituation im Header des JPG-Dateiformats als sogenannte Exif-Information ab. Diese Daten verarbeitet VisualSFM beim Einlesen der Fotos. Die wichtigsten Größen sind die Abmessung des Fotos in Pixel und die eingestellte Kamerabrennweite. Ist letzteres nicht vorhanden, so berechnet VisualSFM diese nach einer festgelegten Formel.

$$\text{Brennweite [px]} = 1.2 * \max(\text{Bildbreite[px]}, \text{Bildhöhe[px]}) \quad (1)$$

Die berechnete Brennweite ergibt sich aus dem Faktor 1.2 multipliziert mit der größeren Kantenlänge des Fotos. Das Ergebnis dieser Formel ist bei den verwendeten Kameras viel zu groß und führt zu deformierten 3D-Modellen. Sollten die Brennweite in den Exif-Informationen nicht enthalten sein, so ist es unbedingt empfehlenswert, diese nachträglich hinzuzufügen. Da die Brennweite während der Berechnung in jedem Fall angepasst wird, ist eine gute Schätzung vorteilhaft.

## 2.5 CMVS/PMVS2 und CPMVS

Das Ergebnis von VisualSFM ist eine dünne Punktwolke. Für die Verdichtung der Punktwolke sind zwei Wege vorgesehen. Der schnelle und einfache Weg führt über CMVS/PMVS2 (Furukawa & Ponce, 2011). Wurden die Programme vom Anwender manuell zum VisualSFM Verzeichnis hinzugefügt, werden diese direkt über das Menü gestartet. Das Ergebnis von CMVS/PMVS2 ist ein einfaches 3D-Modell, anhand dessen der Erfolg der Auswertung beurteilt werden kann, dem es aber deutlich an Schärfe fehlt (vgl. Abb. 2 links). Die Berechnung des Modells aus 76 Fotos dauerte insgesamt weniger als 30 Minuten. Der längere und aufwändigere Weg führt über das externe Programm CPMVS (Jancosek & Pajdla, 2012). Das Programm benötigt ein Microsoft Windows 7 64Bit, eine CUDA fähige Grafikkarte der Marke NVIDIA und wird über die Eingabeaufforderung gestartet. Für die Berechnung des Modells in Abb. 2 (rechts) benötigte es ca. 6 Stunden. Im Vergleich zu Abb. 2 (links) ist dieses 3D-Modell bedeutend detailreicher (vgl. Abb. 3).



Abb. 2: 3D-Modell mit CMVS/PMVS2 (links) und CPMVS (rechts)



Abb. 3: Vergrößerter Ausschnitt aus Abb. 2 rechts

### 3 123D-Catch

Die Anwendung 123D-Catch hat eine einfach aufgebaute Programmoberfläche. Die Fotos werden in die Anwendung importiert und anschließend an den Autodesk-Server gesendet. Dieser verarbeitet die Fotos und berechnet ein 3D-Modell, das nach Abschluss heruntergeladen werden kann. Die Betrachtung erfolgt ebenfalls in 123D-Catch. Das Modell wird standardmäßig geglättet, wodurch Details verloren gehen und das Modell an Schärfe verliert.

Ein Vorteil von 123D-Catch sind die geringen Anforderungen an die Leistung des Anwender PCs sowie die Tatsache, dass es praktisch keine Einarbeitungszeit erfordert. Da die Daten bereits im Internet liegen, sind sie entsprechend auch andernorts verfügbar. Ein großer Nachteil von 123D-Catch ist der Zwang einer ständigen Online-Verbindung. Selbst wenn man das berechnete Modell auf den Arbeitsrechner geladen hat, ist für dessen Betrachtung mit 123D-Catch eine Internetverbindung erforderlich. Um unabhängig von dieser Software zu werden, ist es daher empfehlenswert, das Modell in ein anderes Format, bspw. \*.ply, zu exportieren.

### 3.1 Auflösung der Fotos

Auf Grund der langsamen Übertragung ist eine Reduzierung der Fotos in ihrer Auflösung verlockend. Die Untersuchungen führten zu keinem eindeutigen Ergebnis. Teils wurden mit den reduzierten Fotos besser Ergebnisse erzielt als mit den Originalen (vgl. Abb. 4).

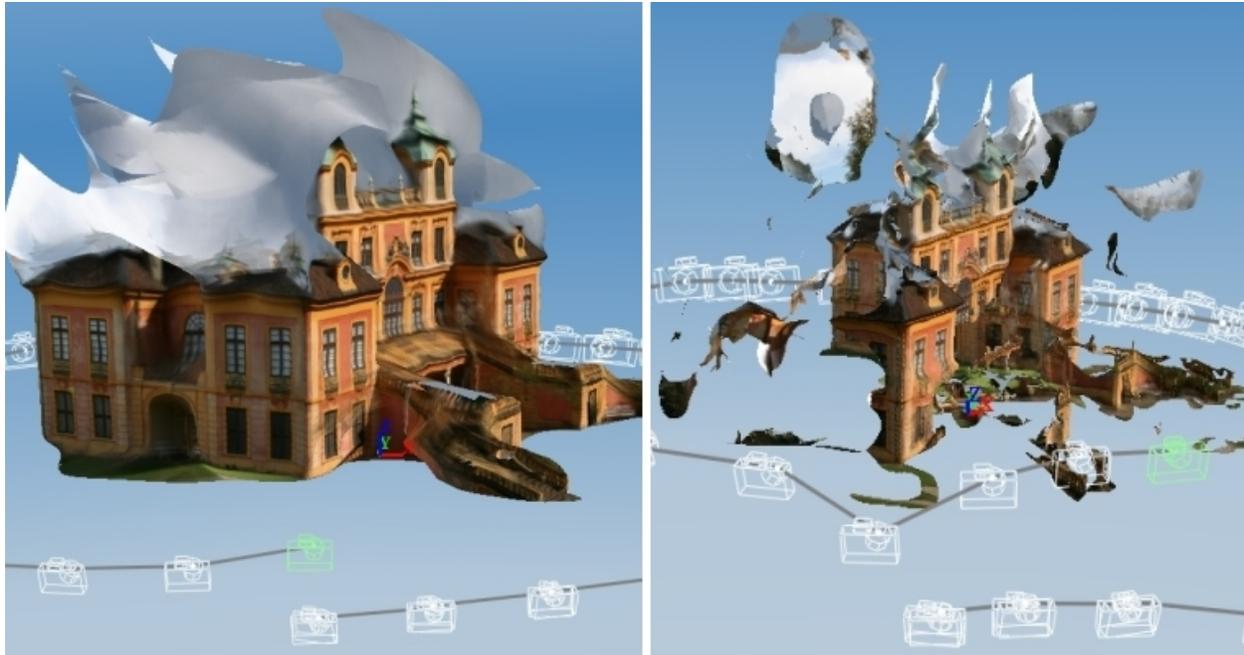


Abb. 4: 123D-Catch Modell aus reduzierten Fotos (links) und originalen Fotos (rechts)

### 3.2 Exif-Informationen

Versendet man Bilder an den Autodesk-Server, bei denen keine Exif-Informationen vorhanden sind oder diese vorher entfernt wurden, kommt es zu erheblichen Deformationen. Es ist dabei unerheblich, ob die Originalaufnahmen oder die reduzierten Fotos versendet wurden. Die Exif-Informationen sollten daher unbedingt erhalten bleiben, auch bei einer Reduzierung. Sind zu den Fotos keine Exif-Informationen vorhanden, ist es erforderlich, diese mit entsprechenden Programmen nachträglich hinzuzufügen.

### 3.3 Manuelle Verknüpfung

123D-Catch bietet über das Menü die Möglichkeit, das 3D-Modell durch manuelles setzen von Verknüpfungspunkten zu verbessern. In dem halbautomatischen Verfahren kann der Anwender pro Bild vier Punkte setzen und diese in benachbarte Bilder übertragen. Die Einfärbung der gesetzten Punkte signalisiert dem Anwender, ob 123D-Catch die Verknüpfung herstellen konnte. Getestet wurde diese manuelle Nachbearbeitung an einem deformierten Modell (vgl. Abb. 5 links). Trotz ausgiebiger Bearbeitung konnte in diesem Fall kein zufriedenstellendes Ergebnis erzielt werden (vgl. Abb. 5 rechts).

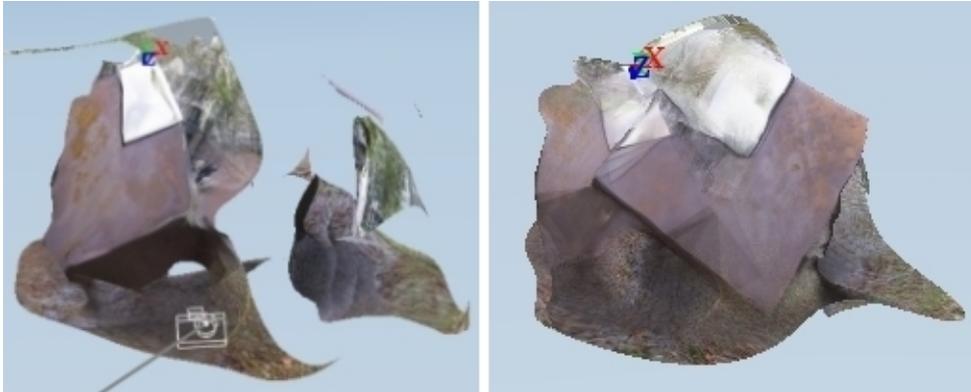


Abb. 5: Ergebnis von 123D-Catch, vor (li.) und nach (re.) der manuellen Nachbearbeitung

## 4 Ausgewählte Ergebnisse und Vergleiche

VisualSFM überzeugt mit guten Ergebnissen bei unterschiedlichsten Anforderungen.

### 4.1 Beispiel Gebäude

Am Beispiel eines Gebäudes mit sehr vielen Kleinstrukturen wurde der Vergleich der beiden Verfahren durchgeführt. In Abb. 6 ist sehr deutlich zu sehen, dass VisualSFM in Kombination mit CMPMVS die deutlich besseren Ergebnisse liefert als 123D-Catch.



Abb. 6: 3D-Modell berechnet mit VisualSFM & CMPMVS (links) und 123D-Catch (rechts)

### 4.2 Beispiel Würfel

Abb. 7 zeigt vier 3D-Modelle eines Würfels mit bekannten äußeren Maßen und z.T. sehr homogener Oberfläche. Die beiden linken Modelle aus den Fotos der Canon G12 unterscheiden sich kaum, wobei einmal reduzierte und einmal die originalen Fotos für die Berechnung verwendet wurden. Die beiden rechten Modelle aus den Fotos der Canon 400D weisen sowohl bei den reduzierten als auch bei den originalen Fotos Deformationen auf. Gründe hierfür sind äußere Einflüsse wie das Tageslicht und Unterschiede bei den Aufnahmeparametern wie z.B. die Aufnahmeentfernung, die Überlappung der Fotos oder der Kameraeinstellungen.

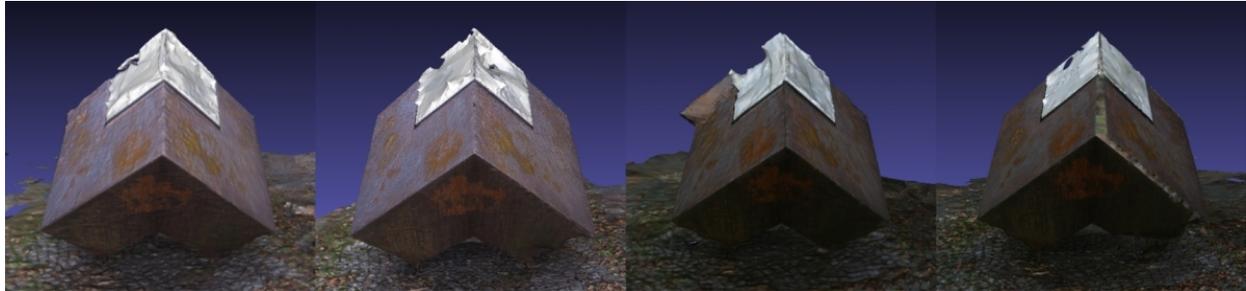


Abb. 7: 3D-Modelle eines Würfels berechnet mit VisualSFM & CPMVS (li. nach re.: G12 reduzierte Fotos, G12 originale Fotos, 400D reduzierte Fotos, 400D originale Fotos)

### 4.3 Beispiel Fassade

Auch sehr feine Details zum Bauwerksmonitoring können mit VisualSFM und CPMVS erfolgreich modelliert werden. Abb. 8 zeigt links das texturierte 3D-Modell des Fassadenausschnittes, in der Mitte den gleichen Ausschnitt ohne Textur. Der Ausschnitt beträgt ca. 30x50cm, wobei die vertikale Fuge ca. 37cm lang ist. Abb. 8 rechts zeigt eines der fünf Fotos des Datensatzes die mit einer Aufnahmeentfernung von ca. 1m um die Fuge gemacht wurden. Hier bietet die Photogrammetrie durch Aufnahmen auch in geringer Distanz eine gute Alternative zur Aufnahme mit terrestrischem Laserscanning.



Abb. 8: Fassade der HFT, links 3D-Model texturiert, Mitte 3D-Model ohne Textur, rechts Farbbild des Datensatzes

## 5 Fazit und Ausblick

VisualSFM hat sich als sehr robustes und leistungsfähiges Programm herausgestellt. Im Vergleich zu 123D-Catch bietet VisualSFM mehr Einstellungsmöglichkeiten und kommt schneller zu einem ersten Ergebnis. Die 3D-Modelle sind detaillierter und wirken schärfer.

Sowohl bei VisualSFM als auch bei 123D-Catch ist es sehr wichtig, dass die korrekte Brennweite, mit der die Fotos aufgenommen wurden, in den Exif-Informationen enthalten ist.

Fehlende Exif-Informationen oder eine zu starke Abweichung zur tatsächlichen Brennweite wirken sich negativ auf das Ergebnis aus. Diese erstrecken sich von leichten Deformationen bis zu fehlgeschlagenen Modellen. VisualSFM ermöglicht dem Anwender großen Einfluss auf den Auswertungsprozess zu nehmen. Dies sollte genutzt werden, um die Auswertung zu beschleunigen oder fehlerhafte Berechnungen zu korrigieren. Die Entzerrung der Fotos durch VisualSFM ist ausschließlich radialsymmetrisch und im Normalfall völlig ausreichend. Für die Verdichtung der Punktwolke sind die Programme CMVS/PMVS2 oder CMPMVS vorgesehen. Ersteres eignet sich gut für einen „ersten Blick“, da die Berechnungszeit vergleichsweise kurz ist. Mit CMPMVS kann eine wesentlich bessere Modellqualität erreicht werden. Dies erfordert aber auch wesentlich mehr Berechnungszeit.

Da bei VisualSFM die Berechnungen lokal stattfinden, hängt die Leistungsfähigkeit des Computers unmittelbar mit der Berechnungszeit zusammen. Bei 123D-Catch stehen die Hardwareressourcen im Hintergrund, dafür ist allerdings fast ausnahmslos eine Internetverbindung erforderlich.

VisualSFM in Verbindung mit CMPMVS bringt die Photogrammetrie einen Schritt weiter sich als Werkzeug für die zügige und kostengünstige Erstellung detailreicher 3D-Modelle zu etablieren. Die hohe Verfügbarkeit von Kameras ermöglicht eine breite Anwenderschicht und damit Potential für zukünftige Geschäftsfelder.

## 6 Literaturverzeichnis

- Albertz, J., & Wiggenhagen, M. (2009). *Taschenbuch zur Photogrammetrie und Fernerkundung, 5. Aufl.* Heidelberg/ München/ Landsberg/ Frechen/ Hamburg: Wichmann.
- Autodesk. (2012). *123D-Catch*. Abgerufen am 1. Oktober 2012 von <http://www.123dapp.com/catch#catchApps>
- Furukawa, Y., & Ponce, J. (2011). *CMVS-PMVS2*. Abgerufen am 1. Oktober 2012 von <https://github.com/TheFrenchLeaf/CMVS-PMVS>
- Jancosek, M., & Pajdla, T. (2012). *CMPMVS 0.6.0*. Abgerufen am 23. November 2012 von <http://ptak.felk.cvut.cz/sfmservice/websfm.pl?menu=cmpmvs>
- Lowe, D. G. (1999). *Object Recognition from Local Scale-Invariant Features*. Abgerufen am 3. Februar 2014 von <http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=790410>
- NVIDIA GmbH. (2012). *CUDA Toolkit 4.2*. Abgerufen am 1. Oktober 2012 von <http://developer.nvidia.com/cuda/cuda-downloads>
- Wu, C. (2007). *SiftGPU Implementation of Scale Invariant Feature Transform*. Abgerufen am 23. Januar 2014 von <http://cs.unc.edu/~ccwu/siftgpu>
- Wu, C. (2012). *VisualSFM 0.5.21*. Abgerufen am 26. September 2012 von <http://homes.cs.washington.edu/~ccwu/vsfm/>