

# Kamera-Navigation in Punktwolken

Jan BÖHM

## 1 Zusammenfassung

Dieser Beitrag behandelt eine Methode zur automatischen Bestimmung der äußeren Orientierung für jedes einzelne Bild aus einer Bildsequenz, wobei das Vorwissen über die Szenerie in Form eines punkt-basierten Umgebungsmodells vorliegt. Zur Orientierung der Bilder werden Landmarken über die Bildsequenz verfolgt, die mit einer kalibrierten Kamera aufgenommen wird. Als Landmarken werden Intensitätsmerkmale verwendet, die vorher aus dem punkt-basierten Umgebungsmodell durch automatische Merkmalsextraktion abgeleitet wurden.

## 2 Einleitung

Dieser Beitrag führt ein punkt-basiertes Umgebungsmodell (PEM – Point-based Environment Model) ein, welches ein übergeordnetes Bezugssystem repräsentiert und das Vorwissen über die Szenerie vorhält. Das PEM entsteht durch eine dichte punkt-weise Abtastung der Oberfläche aller Objekte in der Szene. Jede Abtastung liefert die dreidimensionalen Koordinaten des jeweiligen Punktes der Oberfläche, sowie einen Helligkeitswert, der die Reflektivität der Oberfläche widerspiegelt.

Die Motivation für ein solches Umgebungsmodell entsteht aus der Erwartung, dass in überschaubaren Zeiträumen dichte Punktwolken von größeren öffentlichen Gebäudekomplexen, wie Flughäfen und Bahnhöfen, von Fabrikanlagen und von ganzen urbanen Gebieten verbreitet verfügbar sein werden. Dies wird möglich durch den zunehmenden Reifegrad der am Markt verfügbaren Sensorsysteme, also im Wesentlichen der Laserscanner und durch einen wachsenden Markt an Dienstleistern. Die hauptsächlichen Impulse für die verbreitete Erfassung kommen derzeit aus dem Computer Aided Facility Management und der Fabrikplanung. Sind die Daten allerdings einmal erfasst, kann hieraus eine Vielzahl von Produkten beziehungsweise Anwendungen abgeleitet werden, die zum Teil weit über die ursprüngliche Zielsetzung hinaus gehen.

Da das computergestützte Bildverstehen nach wie vor ein sehr schwieriges Problem ist, ist es nicht überraschend, dass auch die automatische Modellierung von Punktwolken große Probleme bereitet. Aufgrund des Dilemmas, dass einerseits Punktwolken in großem Umfang erfasst werden können und andererseits die Modellierung aufgrund mangelnder Automation einen Flaschenhals in der Verarbeitungskette darstellt, werden Verfahren interessant, die auf eine explizite Modellierung verzichten. Beispiele für derartige Verfahren sind die Übermaß- beziehungsweise Kollisionserkennung von im CAD geplanten Bauteilen, zum Beispiel Rohrleitungen, mit Laserscanning Rohdaten, die die gebaute Realität abbilden. Auch in der punkt-basierten Computergrafik wird die explizite Modellierung vermieden und zum Beispiel durch dynamische Berechnungen der Nachbarschaftsbeziehungen

ersetzt. Diese Verfahren können auch zur effizienten Visualisierung von Laserscanning Daten eingesetzt werden (BÖHM & PATERAKI, 2006).

Auch dieser Beitrag thematisiert ein Verfahren welches direkt, also ohne vorherige Modellierung, auf der Punktwolke aufsetzt. Die zum Ziel gesetzte Anwendung ist dabei die kamera-basierte Navigation. Das PEM liefert dabei die Landmarken, die für die Orientierung der Kamera benötigt werden. Als Landmarken werden Merkmalspunkte verwendet, die in der Bildsequenz, welche die Kamera liefert, verfolgt werden. Der Kern des Orientierungsverfahrens beruht dabei auf bekannten Vorarbeiten aus dem Bereich Photogrammetrie und Computer Vision, die im Folgenden kurz zusammengefasst werden sollen.

### 3 Vorarbeiten

Die Orientierung eines einzelnen Bildes, bei der die Drehung und Verschiebung zu einem übergeordneten Bezugssystem berechnet wird, ist ein zentrales und gut untersuchtes Problem in der Photogrammetrie. Auch die Orientierung von Bildern aus Bildsequenzen ist eine bekannte Aufgabenstellung, deren vielfältige Anwendungen zum Beispiel die Navigation, Robotik, Augmented Reality oder auch die Objekterkennung umfassen. Die Lösungsansätze für diese Aufgabenstellung entstammen zumeist dem Bereich des „Structure-from-Motion“ (SfM) in Kombination mit einem geeigneten Operator zur Extraktion und Verfolgung von Merkmalen in Bildern (POLLEFEYS, KOCH & GOOL, 1998; TOMASI & KANADE, 1992). SfM benötigt kein Vorwissen über die Szene. Daher stehen zunächst auch keine Landmarken zur Orientierung in einem übergeordneten Bezugssystem zur Verfügung. Stattdessen werden beliebige geeignete Merkmalspunkte zwischen den einzelnen Bildern einer Sequenz verfolgt. Die Orientierung der Bilder beruht dann auf dem Verfahren der relativen Orientierung, sofern die Kamera kalibriert ist. Für unkalibrierte Kameras kann auch ein Fundamentalmatrix-Ansatz verwendet werden. Anders als in der für diesen Beitrag formulierten Zielsetzung berechnet SfM nicht nur die Orientierungen der Kamera, sondern versucht gleichzeitig eine Rekonstruktion der Szene. Diese Berechnungen erfolgen bei den oben beschriebenen Eingangsinformationen lediglich in einem Modellkoordinatensystem.

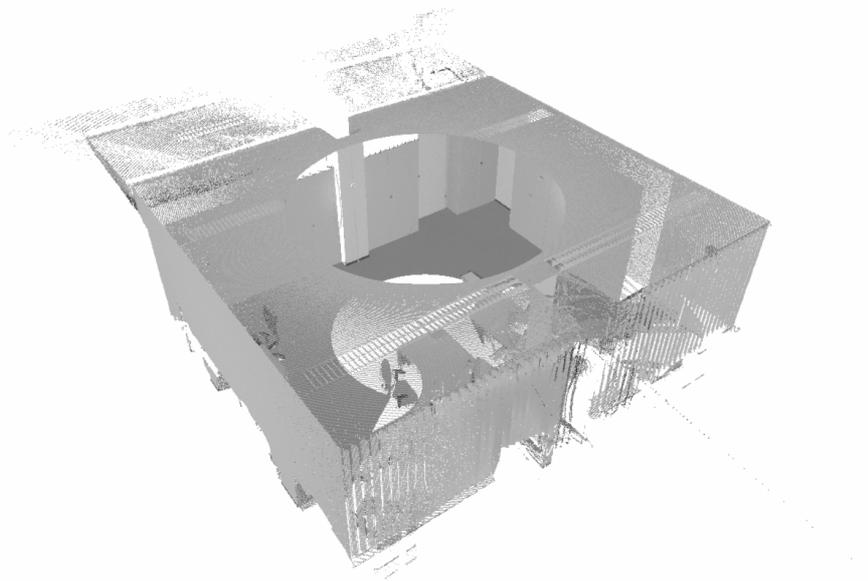
Die Berücksichtigung von Vorwissen in Form von Distanzdaten für derartige Aufgabenstellungen hat in jüngster Zeit wieder verstärkt Interesse gefunden. ZHAO u. a. (2005) schlagen ein Verfahren vor, welches SfM mit dem „Iterative Closest Point“ (ICP) Algorithmus verbindet. Zunächst wird einfaches SfM eingesetzt. Die daraus errechnete Punktwolke wird dann mit den gegebenen Distanzdaten mithilfe des ICP Algorithmus in Übereinstimmung gebracht. Dieses Verfahren wurde erfolgreich auf Luftbildsequenzen getestet, wobei allerdings zusätzliche Orientierungssensoren eingesetzt wurden.

LERNER u. a. (2006) schlagen ein alternatives Verfahren vor, bei dem das Vorwissen in Form eines Digitalen Höhenmodells (DHM) vorliegt. Das DHM wird verwendet, um die Komplexität der Orientierungsberechnung zu verringern. Dabei werden die Bildstrahlen an der Oberfläche, die das DHM repräsentiert, abgeschnitten. Dies vermeidet eine explizite Rekonstruktion der Oberfläche. Auch dieses Verfahren wurde auf einer Luftbildsequenz und zusätzlichen Informationen über die Bewegungsbahn getestet.

## 4 Punkt-basiertes Umgebungsmodell

Das PEM und die daraus abgeleiteten Landmarken dienen als Bezugssystem für die anschließende Bildorientierung. Das PEM besteht im Wesentlichen aus einer dichten Punktwolke mit assoziierten Intensitätswerten. Es existieren viele Ansätze und Verfahren zur Erfassung dichter Punktwolken. Sowohl passive als auch aktive Triangulationsverfahren und Time-of-Flight (TOF) Verfahren werden eingesetzt. Der Vorteil eines TOF Scanners liegt in der Regel in der Robustheit der Erfassung auch über große Distanzen und bei ungünstigen Beleuchtungsverhältnissen. Die Genauigkeit und die Zuverlässigkeit der dreidimensionalen Information der Punktwolke sind elementar wichtig, da sie wesentlich die Genauigkeit der Kameraorientierung bestimmen. In unseren Untersuchungen setzten wir den Leica HDS 3000 für die Datenerfassung ein. Der 360 Grad Scan eines Büroraums, der als Testszenario dient, ist in Abbildung 1 dargestellt.

Das regelmäßige polare Raster, in dem die Punktwolke vom Scanner erfasst wird, erlaubt es die gewonnenen Intensitätswerte als Bild darzustellen. Dabei soll jedoch betont werden, dass die vom Laserscanner erfassten Intensitätswerte grundsätzlich unterschieden werden müssen von Bildern, die mit einer Kamera aufgenommen wurden. Zum einen repräsentieren die Intensitätswerte des Laserscanners die Reflektivität der gemessenen Oberfläche in einem sehr schmalen Wellenlängenbereich (zum Beispiel 532 nm). Zum anderen stimmen beim Laserscanner Beleuchtungsrichtung und Beobachtungsrichtung genau überein, wohingegen bei einer Kamera meist eine externe Lichtquelle, zum Beispiel das Umgebungslicht, die Oberfläche beleuchtet. Daher kommt es in Kamerabildern auch oft zu starken Schattenwürfen, die wiederum unzuverlässige Merkmalspunkte erzeugen können.

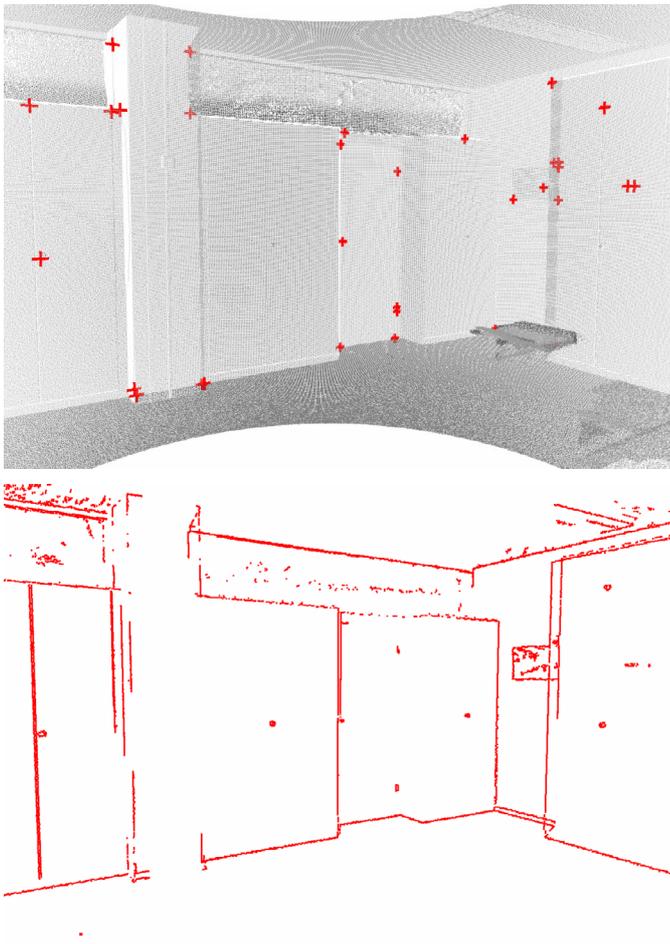


**Abb. 1:** Der 360 Grad Scan eines Büroraums.

#### 4.1 Extraktion von Landmarken

Wenn wir diesen Unterschied in der Bildgewinnung ignorieren, können wir Standard-Bildverarbeitungsalgorithmen einsetzen, um die Intensitätsbilder des Laserscanners zu verarbeiten. Um markante Punkte im Umgebungsmodell zu extrahieren, kann die bekannte „Corner-Response-Function“ von HARRIS & STEPHENS (1988) eingesetzt werden. Das Ergebnis dieser Verarbeitungsstufe ist in Abbildung 2 oben dargestellt. Die extrahierten Landmarken sind als Kreuze in der Punktwolke markiert.

Zur beschleunigten Darstellung des Umgebungsmodells im Bild wird eine reduzierte Variante der Punktwolke errechnet. Diese soll die Geometrie der Szene jedoch immer noch hinreichend beschreiben. Daher wird ein einfaches Kantenextraktionsverfahren eingesetzt, um Punkte auf Kanten aus dem Umgebungsmodell zu extrahieren. Diese Ergebnisse sind in Abbildung 2 unten zu sehen.



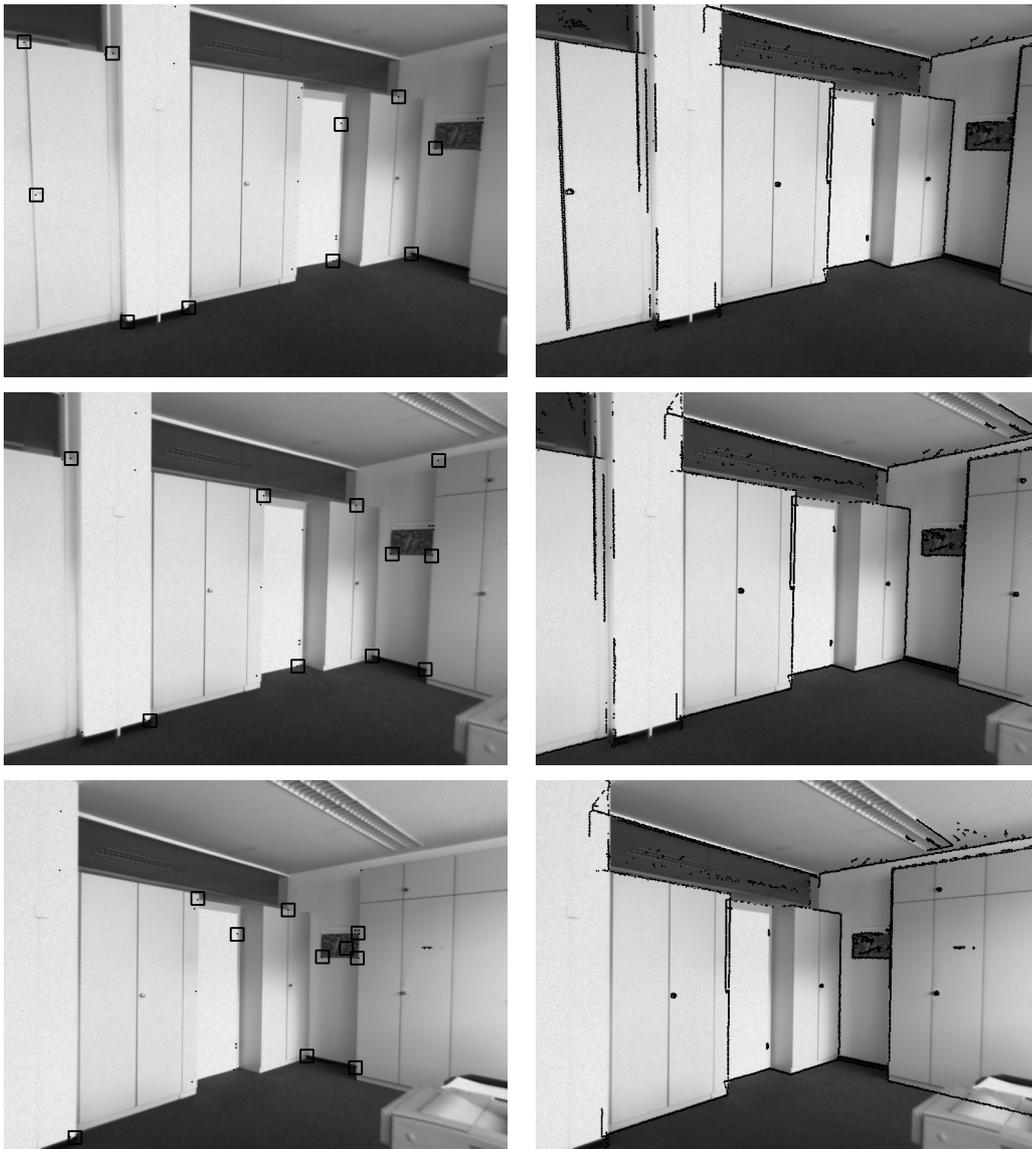
**Abb. 2:** Landmarken, die aus dem Umgebungsmodell extrahiert wurden und eine reduzierte Variante der Punktwolke aus Kantenpunkten.

## 5 Merkmalsverfolgung und Bildorientierung

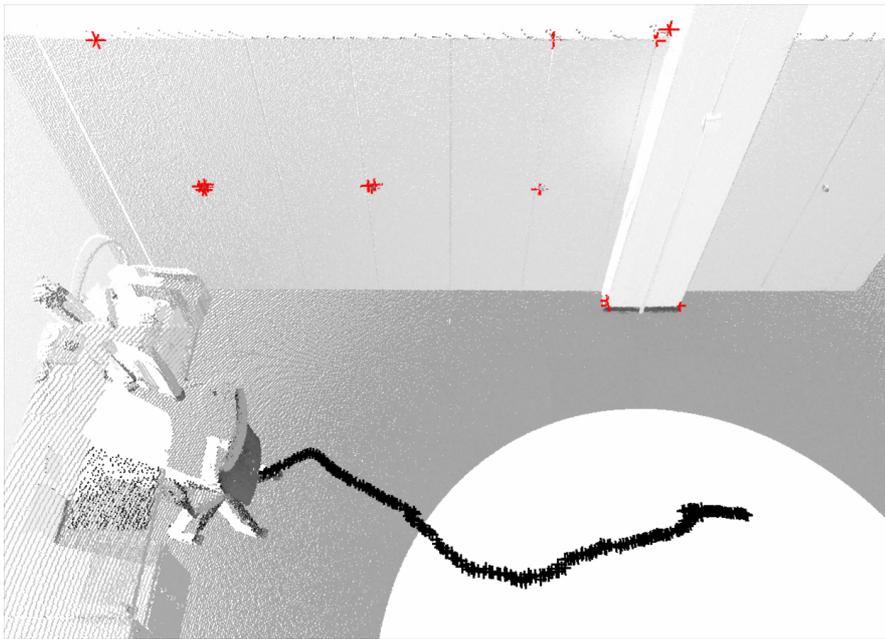
Wie bereits in der Einleitung erwähnt, basiert die Orientierung der Bilder auf der Verfolgung von Merkmalen in der Bildsequenz. Für den Entwurf eines solchen Algorithmus können einige Schlüsselkomponenten identifiziert werden (SHI & TOMASI, 1994). Dazu gehören die Merkmalsextraktion, das Bewegungsmodell, der Zuordnungsalgorithmus und das Ähnlichkeitsmaß. In dem hier beschriebenen Ansatz können keine optimalen Merkmale aus dem Bild extrahiert werden. Stattdessen müssen die vorab aus dem Umgebungsmodell extrahierten Landmarken im Bild verfolgt werden. Dazu wird zufällig eine feste Zahl von Landmarken ausgewählt und in das Bild projiziert. Die Anzahl der Punkte ist dabei ein Kompromiss zwischen Zuverlässigkeit und Bearbeitungsgeschwindigkeit. Um die Landmarken in das Bild projizieren zu können, muss einmalig die Orientierung des ersten Bildes der Sequenz bestimmt werden. Hierzu werden vier ausgewählte Landmarken im Bild manuell identifiziert und die äußere Orientierung des Bildes per räumlicher Rückwärtsschnitt berechnet.

Die Verfolgung der projizierten Landmarken im nächsten Bild einer Sequenz geschieht, indem die absoluten Differenzen in einer Maske um den originalen Punkt und seiner möglichen nächsten Position aufsummiert werden. Diese Berechnungen sind extrem schnell und genügen den Anforderungen, um Merkmale verfolgt zu können. Um falsche Zuordnungen zu eliminieren, wird ein lineares Bewegungsmodell angenommen und der bekannte RANSAC Algorithmus (FISCHLER & BOLLES, 1981) zur Elimination von Ausreißern eingesetzt. Nachdem die Landmarken in das nächste Bild der Sequenz übertragen wurden, kann für dieses Bild wiederum der räumliche Rückwärtsschnitt berechnet werden. Werden einzelne Landmarken nicht mehr detektiert, müssen sie ersetzt werden, um die Anzahl der Punkte nicht degenerieren zu lassen. Dasselbe gilt für Punkte, die aus dem Gesichtsfeld wandern.

In Abbildung 3 ist die erfolgreiche Verfolgung von Landmarken als Merkmalspunkte im Bild dargestellt. Die Berechnungen erfolgten auf einer Bildsequenz von 500 Bildern einer Basler A302f Machine Vision Kamera. Die Kamera verfügt über 780x582 Pixel bei einer Bildrate von 30 Hz. Zusammen mit den Berechnungen und der Darstellung erzielt das System derzeit eine Wiederholrate von 10 Hz. Nachdem für jedes Bild der Rückwärtsschnitt berechnet wurde, kann auch die reduzierte Darstellung der Punktwolke in das Bild projiziert werden. Aus der Überlagerung der dreidimensionalen Kantenpunkte mit den entzerrten Bildern der Kamera kann die Güte der Bildorientierung abgeschätzt werden. Dies ist ebenfalls in Abbildung 3 dargestellt. Abbildung 4 zeigt die Trajektorie der mittels Rückwärtsschnitt berechneten Kamerastandpunkte aus einer Sequenz mit 400 Bildern.



**Abb. 3:** Drei Bilder aus einer Sequenz von 500 Bildern von oben nach unten angeordnet, die eine Bewegung der Kamera von links nach rechts zeigen. In der linken Spalte sind die projizierten Landmarken durch kleine Punkte und die tatsächlich verfolgten Merkmale durch Kästchen dargestellt. In der rechten Spalte sind die Kantenpunkte des Umgebungsmodells mit den Bildern überlagert.



**Abb. 4:** Trajektorie des Projektionszentrums der Kamera gemeinsam mit der Punktwolke und den Landmarken dargestellt.

## 6 Zusammenfassung

Der Beitrag demonstriert erfolgreich die direkte Verwendung einer per Laserscanning aufgenommenen Punktwolke als Referenzsystem zur kamera-basierten Navigation. Bekannte Verfahren der Bildverarbeitung können auf die Intensitätswerte, die während des Scanvorgangs erfasst werden, angewandt werden, um robuste Landmarken zu extrahieren. Die Bildsequenzen einer Videokamera können im gewählten Testszenario zuverlässig im Umgebungsmodell orientiert werden.

## 7 Literatur

Böhm, J. & Pateraki, M. (2006): From Point Samples to Surfaces: On Meshing and Alternatives ISPRS Image Engineering and Vision Metrology, XXXVI, 50-55.

Fischler, M. A. & Bolles, R. C. (1981): Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography, Communications of the ACM, 24(6), pp. 381–393.

R. M. Haralick, C.-N. Lee, K. Ottenberg, and M. Nöelle (1994): Review and analysis of solutions of the three point perspective pose estimation problem, *Int. J. Comput. Vision* 13(3), pp. 331–356, 1994.

Harris, C. & Stephens, M. (1988): A combined corner and edge detector, *Fourth Alvey Vision Conference*, 147-151.

Lerner, R.; Rivlin, E. & Rotstein, H.P. (2006): Pose and Motion Recovery from Feature Correspondences and a Digital Terrain Map, *IEEE Trans. Pattern Anal. Mach. Intell.*, IEEE Computer Society, 28, 1404-1417.

Pollefeys, M.; Koch, R. & Gool, L.V. (1998): Self-Calibration and Metric Reconstruction in spite of Varying and Unknown Internal Camera Parameters, *ICCV '98: Proceedings of the Sixth International Conference on Computer Vision*, IEEE Computer Society, 90.

Tomasi, C. & Kanade, T. (1992): Shape and motion from image streams under orthography: a factorization method, *Int. J. Comput. Vision*, Kluwer Academic Publishers, 9, 137-154.

Shi, J. & Tomasi, C. (1994): Good Features to Track, *IEEE Conference on Computer Vision and Pattern Recognition (CVPR'94)*.

Zhao, W.; Nister, D. & Hsu, S. (2005): Alignment of Continuous Video onto 3D Point Clouds *IEEE Trans. Pattern Anal. Mach. Intell.*, IEEE Computer Society, 27, 1305-1318