

Anwendungen im Umfeld von Panoramabildern

Seminararbeit

vorgelegt am
Lehrstuhl für Praktische Informatik IV
Prof. Dr.-Ing. Wolfgang Effelsberg
UNIVERSITÄT MANNHEIM

von
Dipl. Wirtsch.-Inf. Eric Lunkenbein
aus Lampertheim

Frühjahrssemester 2008

Inhaltsverzeichnis

1	Einleitung und Aufbau der Arbeit	1
2	Anwendungen im Umfeld von Panoramabildern	2
2.1	Hugin	2
2.2	QuickTime VR	4
2.3	Video Indexing	7
3	Moderne Anwendungen im Entwicklungsstadium	14

Abbildungsverzeichnis

2.1	Arbeitsschritte Hugin	4
2.2	Kamerarotation	5
2.3	QuickTime VR File Format	6
2.4	QuickTime VR Panoramic Player	7
2.5	QuickTime VR Authoring	8
2.6	Szenen Einzelbild 1	9
2.7	Szenen Einzelbild 2	10
2.8	Szenen Einzelbild 3	10
2.9	Statische Zusammenfassung	11
2.10	Dynamische Zusammenfassung	11
2.11	Indizierung: Auswahl eines Punktes	12
2.12	Vererbung der Information in alle Einzelbilder	12
3.1	Photosynth	15

Abkürzungsverzeichnis

3D	dreidimensional.
DVI	Digital Video Interactive.
Exif	Exchangeable Image File Format.
GPS	Global Positioning System.
RANSAC	Random Sample Consensus.
SIFT	Scale-invariant feature transform.
WWW	World Wide Web.

1 Einleitung und Aufbau der Arbeit

Bereits vor der Erfindung der Fotografie, gab es Bemühungen die Wahrnehmung einer ganzen Umgebung für spätere Betrachter zu konservieren. Unter Verwendung von runden Gebäuden als *Leinwand* wurde mit Pinsel und Farbe die komplette Rundumsicht einer Landschaft für den Betrachter festgehalten.¹ Nach der Erfindung der Fotografie wurden immer wieder spezielle Kameras und Vorrichtungen entwickelt, um die Erzeugung von Panoramen zu ermöglichen.² Mit der Entwicklung der digitalen Fotografie bzw. Bildverarbeitung haben sich die Möglichkeiten zur Erzeugung und Darstellung von Panoramen enorm vergrößert.

Diese Arbeit stellt ausgewählte Anwendungen für die Erzeugung und Darstellung von Panoramen vor.

- Dabei wird zunächst in Kapitel 2.1 kurz eine Sammlung freier Anwendungsprogramme, die insbesondere die Erstellung von Panoramabildern aus einer Reihe einzelner Aufnahmen eines Motivs ermöglicht, vorgestellt.
- In Kapitel 2.2 folgt die Vorstellung von QuickTime VR, einer Erweiterung des Multimedia Frameworks QuickTime von Apple.
- Die Vorstellung eines Verfahren zur Indizierung von Videos auf Basis von Mosaikbildern folgt in Kapitel 2.3.
- Der abschließende Ausblick in Kapitel 3 geht auf eine moderne Anwendung ein, die sich noch im Entwicklungsstadium befindet. Die Anwendung Photosynth ist in der Lage einzelne Bilder einer gemeinsamen Szene oder Umgebung in ihrem räumlichen Zusammenhang anzuordnen, darzustellen und navigierbar zu machen.

¹vgl.[WWDM99]

²vgl. <http://www.panoramicphoto.com/timeline.htm>

2 Anwendungen im Umfeld von Panoramabildern

Zur Darstellung ausgedehnter Motive oder ganzer Umgebungen eignen sich einzelne herkömmliche Fotografien nur bedingt. Mittels eines Panoramas dagegen lassen sich große Objekte oder ganze Umgebungen in hoher Qualität als Ganzes darstellen.

Bei der Erstellung von Panoramen ist zwischen zwei unterschiedlichen Vorgehensweisen zu unterscheiden, nämlich

- einerseits Methoden unter Verwendung von spezieller Hardware, die ein einziges Bild erzeugen
- andererseits Methoden, die mehrere Einzelbilder zur Erstellung eines Panoramas verwenden.

Im Folgenden werden einige Anwendungen vorgestellt, die im Wesentlichen auf der Verwendung mehrerer Einzelbilder basieren.

2.1 Hugin

Hugin dient als grafische Benutzeroberfläche für eine Reihe von Kommandozeilenanwendungen zur Erzeugung von Panoramabildern aus einer Menge von zusammengehörigen Einzelbildern. Unter Einhaltung der folgenden Bedingungen erlaubt die Verwendung dieser frei verfügbaren Softwaresammlung die Erstellung von qualitativ hochwertigen Panoramabildern ohne den Einsatz von spezieller Hardware oder kostenpflichtiger Software, bei relativ geringem manuellem Aufwand:

- Alle für das Panorama benötigten Teilaufnahmen wurden von weitgehend einem gemeinsamen Punkt erstellt.

- Die Teilbilder müssen jeweils eine Überlappung von etwa 20-30% des Bildinhaltes aufweisen
- Verwendung von möglichst gleichbleibenden Kameraparametern (z.B. Belichtungszeit, Blende, Brennweite, Fokus, Weißabgleich, Objektiv, Blitznutzung)

Die Softwaresammlung die über die gemeinsame Benutzeroberfläche *Hugin*³ gesteuert wird besteht aus mehreren, auch austauschbaren, Komponenten. Üblich ist die Verwendung von folgenden Einzelanwendungen:

- *Autopano*⁴ zur Feststellung identischer Punkte auf den Einzelbildern
- *Enblend*⁵ zur Überblendung der Schnittbereiche
- *Panorama Tools*⁶ mit u.a. *PTStitcher* für das Zusammenfügen der Einzelbilder und *PTOptimizer* für die Bestimmung der Bild-Parameter

Die einzelnen Komponenten sind austauschbar, so lässt sich an Stelle des Autopano Tools auch das erweiterte Autopano SIFT⁷ verwenden.

Für die Verwendung von Hugin lassen sich im World Wide Web (WWW) zahlreiche Tutorien finden.^{8,9,10}

Die Erstellung eines Panoramabildes aus einer Reihe von Einzelbildern eines Motivs unter Verwendung dieser Tools läuft in mehreren Schritten ab (siehe Abbildung 2.1).

Mit den zuvor beschriebenen Programmen erstellte Panoramabilder können unter Verwendung einer weiteren frei verfügbaren Anwendung (*Pano2QTVR free*)¹¹ in das QuickTime VR-Format konvertiert werden. Das folgende Kapitel stellt dieses Format und die zugehörige Anwendung QuickTime VR vor.

³<http://hugin.sourceforge.net/>

⁴<http://autopano.kolor.com/>

⁵<http://enblend.sourceforge.net/>

⁶<http://panotools.sourceforge.net/>

⁷vgl.[Low04]

⁸<http://hugin.sourceforge.net/tutorials/index.shtml>

⁹http://www.panoclub.de/hugin_tut/

¹⁰http://freenet-homepage.de/martin_wehner/Tutorial/panorama.html

¹¹<http://www.pano2qtv.com/>

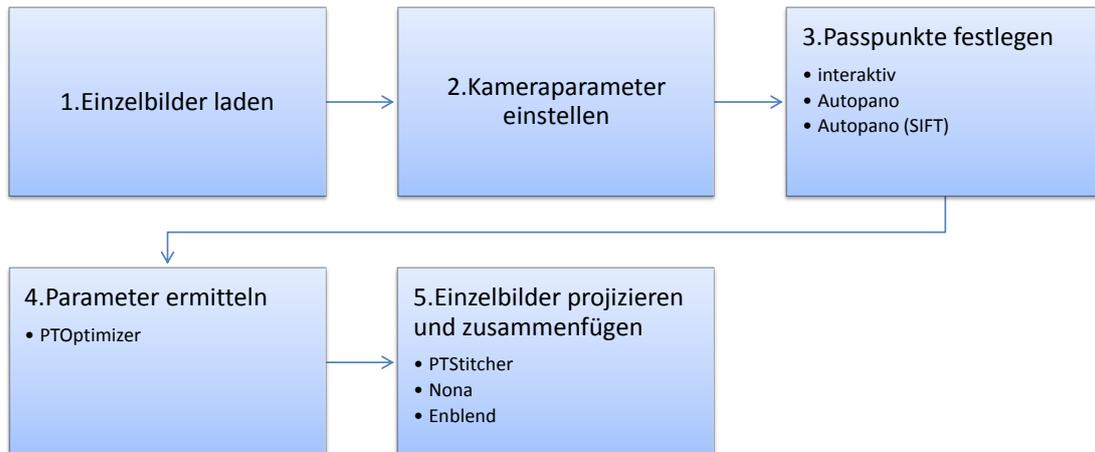


Abbildung 2.1: Arbeitsschritte Hugin

2.2 QuickTime VR

Eines der Hauptziele bei der Erstellung von virtuellen Umgebungen ist es, dem Betrachter soweit wie möglich den Eindruck zu vermitteln, sich tatsächlich in dieser Umgebung zu befinden. Hierzu bedarf es einer detailreichen Darstellung dieser Umgebung. Traditionell wurden virtuelle Umgebungen unter Verwendung von 3D-Computer-Grafiken erstellt. Diese Methode zieht zum Einen den Einsatz spezieller Hardware und zum Anderen ein hohes Maß manueller Bearbeitung nach sich. Um die Umgebungen in Echtzeit wiedergeben zu können, sind die Komplexität und Qualität der Modellierung durch die Hardware beschränkt. Auch die sich rapide entwickelnde Hardware wird den Wunsch nach immer mehr Details und größerer Szenenkomplexität nicht kompensieren können. Ein Ansatz zur Lösung dieser Probleme war die Verwendung von sogenannten *branching movies* - *verzweigten Filmen*. Dabei wurden mehrere Filmsequenzen entlang eines vorgegebenen Weges verbunden. Bei der Wiedergabe waren nur definierte Verzweigungspunkte für die Navigation erlaubt. Frühe Beispiel für die Verwendung dieses Ansatzes ist das Projekt *Movie-Map*¹², sowie eine auf Digital Video Interactive (DVI) basierende Demonstration, die die Maya Ruine von Palenque abbildete¹³.

Der Branching-Movies-Ansatz löst zwar die Komplexitäts- und Qualitätsprobleme der 3D-Modellierung, bietet jedoch nur eingeschränkte Navigationsmöglichkeiten und

¹²vgl. [Lip80]

¹³vgl. [Rip89]

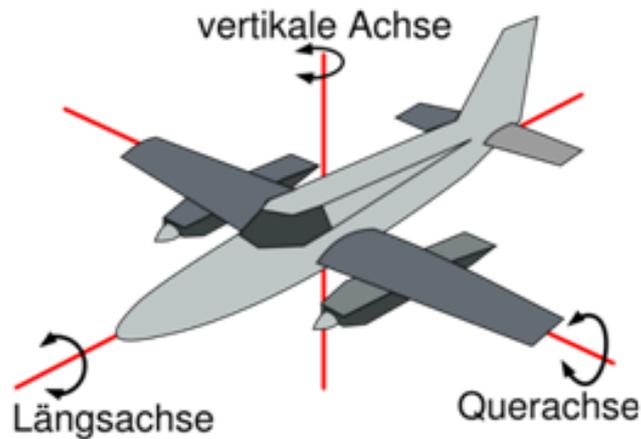


Abbildung 2.2: Kamerarotation Quelle:Wikipedia

ist durch einen sehr hohen Speicherbedarf gekennzeichnet. Aus dieser Motivation hat CHEN¹⁴ im Jahre 1995 einen neuartigen Ansatz zur Darstellung von virtuellen Umgebungen vorgestellt. Dieser Ansatz hat Eingang in das kommerzielle Produkt QuickTime VR, als Erweiterung des QuickTime Multimedia Framework von Apple, gefunden. Als Entwicklungsziele für QuickTime VR waren drei Punkte wesentlich:

- Wiedergabemöglichkeit auf Standard-PC
- Reale und synthetische Umgebungen darstellen
- Hohe Bildqualität ohne Beeinträchtigung der Geschwindigkeit

Diese Entwicklungsvorgaben konnten unter Verwendung der Methode der Wiedergabe von virtuellen Umgebungen auf Basis von Bildern erreicht werden. Wesentliches Element der QuickTime VR Technik ist die Verwendung sogenannter *Environment Maps*¹⁵. Als Environment Map wird die Projektion einer Szene auf eine geometrische Form verstanden. Nach der Erzeugung einer solchen Environment Map zu einer Szene ist, bei fixem Standpunkt, jede beliebige Ansicht der Umgebung darstellbar. Das heißt, die Verwendung von Environment Maps ermöglicht die Abbildung der grundsätzlich sechs Freiheitsgrade der Kamerabewegung mit der einzigen Einschränkung auf einen fixen Standpunkt.

Freiheitsgrad und verwendete Environment Map:

¹⁴[Che95]

¹⁵vgl. [Gre86]

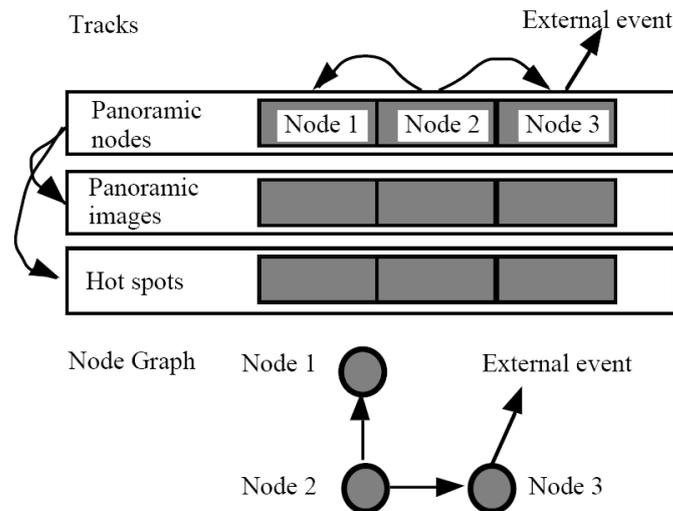


Abbildung 2.3: QuickTime VR File Format Quelle:[Che95] S.4

- 1.-3.: Rotation in 3 Achsen (siehe Abbildung 2.2): Würfel (Rundumsicht), Kugel (Rundumsicht), Zylinder (360° horizontaler Schwenk, <180° Gieren)
- 4.: Objekt-Rotation, Äquivalent zur Drehung der Kamera um das Objekt: Nicht darstellbar mit einzelner Environment Map
- 5.: Kamerabewegung, d.h. Veränderung von Blickrichtung: (einzelne Environment Map) Standpunkt: (Vielzahl von Environment Maps nötig)
- 6.: Veränderung des Blickfeldes (Zoom): Beliebige Environment Map.¹⁶

Zunächst wurden für QuickTime VR ausschließlich zylindrischen Environment Maps eingesetzt.

Für QuickTime VR ist es gelungen, das bestehende Standard QuickTime Dateiformat auch zur Speicherung von virtuellen Umgebungen zu nutzen. Die Umgebung wird als ereignisgesteuerter (im Gegensatz zu zeitgesteuerten klassischen Videos), räumlich orientierter Film in drei Tracks gespeichert (siehe Abbildung 2.3). Zur Wiedergabe wird der *Panoramic Player* verwendet. Bei der Speicherung wird das Panoramabild in einzelne Kacheln zerschnitten. Bei der Wiedergabe werden nun die sichtbaren Teilausschnitte geladen, projiziert und über einen Puffer wiedergegeben (siehe Abbildung

¹⁶Bei Verwendung von Bildvergrößerung ist keine Erhöhung des Detailgrades möglich. Daher Verwendung von mehreren Bildern in unterschiedlicher Auflösung und derselben Blickrichtung: Auswahl der Auflösung nach Zoomlevel möglich.

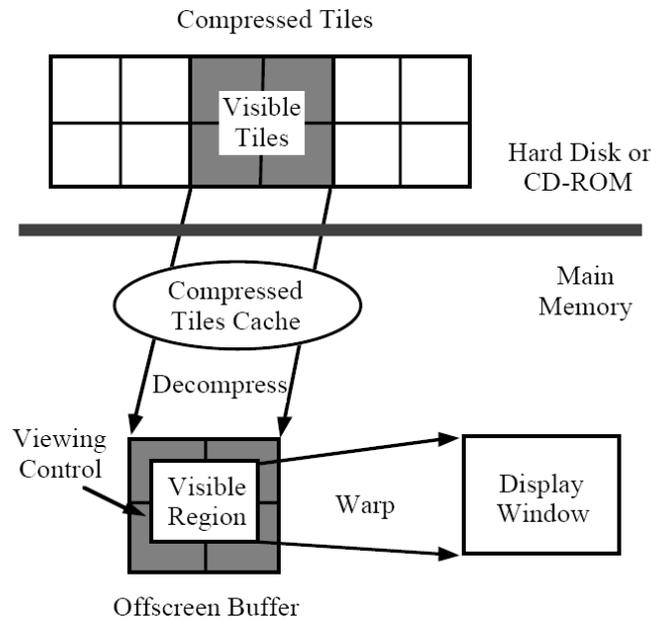


Abbildung 2.4: QuickTime VR Panoramic Player Quelle:[Che95] S.5

2.4). Zur Performancesteigerung können unterschiedliche Bildqualitäten während eines Schwenks verwendet werden, d.h. während der Kamerabewegung wird das Bild in geringer Auflösung wiedergegeben und die volle Auflösung erst beim Kamerastillstand angezeigt. Der Wechsel des Standpunktes, ist nur an definierten Stellen (*sog. Hot Spots*) möglich.

Die Erstellung von QuickTime VR Filmen veranschaulicht Abbildung 2.5. Mit QuickTime VR konnten die eingangs beschriebenen, wesentlichen Entwicklungsziele erreicht werden. Für die Wiedergabe von virtuellen Umgebungen bleiben jedoch zwei wesentliche Einschränkungen erhalten. Zum Einen setzt diese Technik eine statische Szene voraus (lösbar durch zeitabhängige Environment Maps) und zum Anderen ist die Navigation auf exakt definierte Standpunkte limitiert.

2.3 Video Indexing

Video spielt als umfassende Informationsquelle eine enorme Rolle. Allerdings enthalten Videos naturgemäß, bezüglich ihrer semantischen Inhalte, hohe Redundanzen. Durch die chronologische Anordnung einzelner Informationen zu Sequenzen und die Abspielsteuerung alleine über eine Zeitangabe ist kein direkter Zugriff auf bestimm-

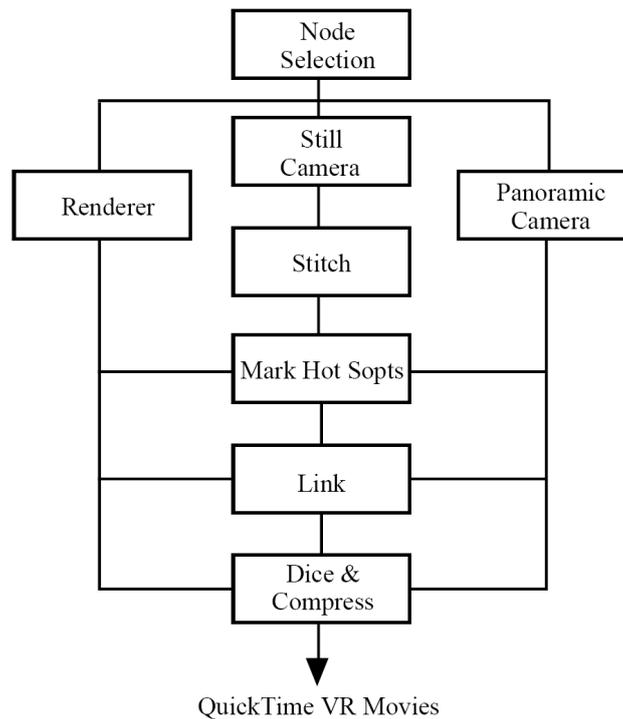


Abbildung 2.5: QuickTime VR Authoring Quelle:[Che95] S.6

te Inhalte möglich. *Irani* und *Anandan*¹⁷ haben 1998 ein Verfahren vorgestellt, um den effizienten Zugriff auf Informationen in Videos und eine effiziente Bearbeitung von Videos zu ermöglichen. Das Verfahren beruht auf der Umwandlung des Videos in mehrere kompakte Sequenz-Darstellungen. Hierzu wird zunächst das Video in Sequenzen unterteilt. Eine Sequenz enthält alle Einzelbilder einer Szene bzw. einer Umgebung im Zeitablauf. Die gemeinsame Information aller Einzelbilder ist somit die Szene oder Umgebung selbst. Die Einzelbildfolge ist für die Wiedergabe als Film geeignet, erlaubt jedoch keinen direkten Zugriff auf Inhalte. Ebenso ist die Bearbeitung nur auf Basis von einzelnen Bildern (sog. *Frames*) möglich. Während die Umgebung die Redundanz innerhalb einer Sequenz ausmacht, sind dennoch etliche Informationen auf viele Einzelbilder verteilt. So ergibt sich die gesamte Umgebung, etwa bei einem Kameranenschwenk, erst durch die Summe aller Bilder. Auch der Informationsgehalt dynamischer Elemente kann nur über die Summe der Einzelbilder erfasst werden. Für die

¹⁷[IA98]

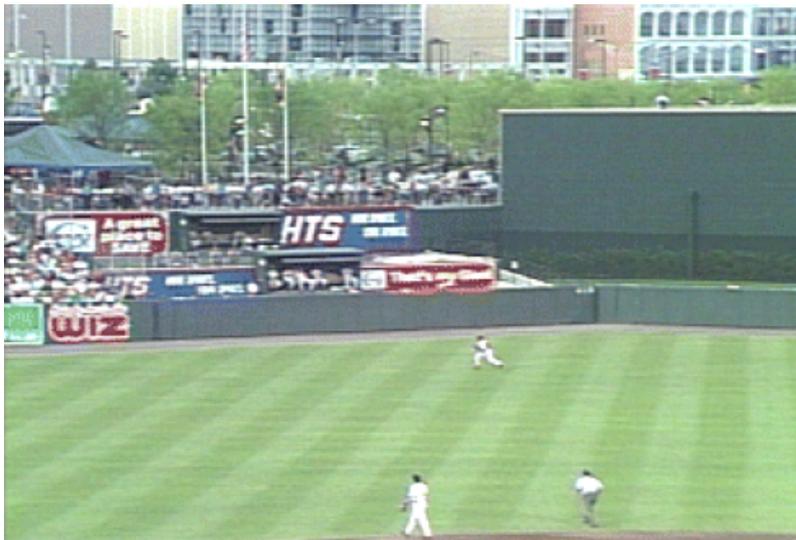


Abbildung 2.6: Szenen Einzelbild 1 Quelle:[IA98] S.5

einzelnen Sequenzen eines Videos wird im Verfahren eine Szenen-Darstellung erzeugt. Diese enthält drei grundlegende Informationsarten:

- **Erweiterte Information zur Umgebung:** Ein Panorama Bild¹⁸ der gesamten Umgebung einer Szene (siehe Abbildung 2.9, als Zusammenfassung der Einzelbilder Abbildung 2.6, Abbildung 2.7 und Abbildung 2.8)
- **Zeitabhängige Information:** Die Bewegung der dynamischen Objekte (siehe Abbildung 2.10)
- **Geometrische Information:** Räumliche Information und die geometrische Transformation aller Objekte

Für bewegte Objekte erfolgt das Erfassen der Bewegungsbahnen über die Zeit. Aus den Einzelbild-Informationen wird auch die Geometrie der Umgebung errechnet und somit der räumliche Zusammenhang bestimmt. Die Einzelbilder erhalten ein gemeinsames Koordinatensystem.

Diese Szenen-Darstellung erlaubt Nicht-Lineares-Navigieren und, über die enthaltenen geometrischen und dynamischen Informationen, ein hoch effizientes Indizieren der Bildinhalte. Dabei können auch Verfahren der klassischen, einzelbildbasierenden

¹⁸vgl. [IAB⁺96]



Abbildung 2.7: Szenen Einzelbild 2 Quelle:[IA98] S.5



Abbildung 2.8: Szenen Einzelbild 3 Quelle:[IA98] S.5



Abbildung 2.9: Statische Zusammenfassung Quelle:[IA98] S.5



Abbildung 2.10: Dynamische Zusammenfassung Quelle:[IA98] S.5

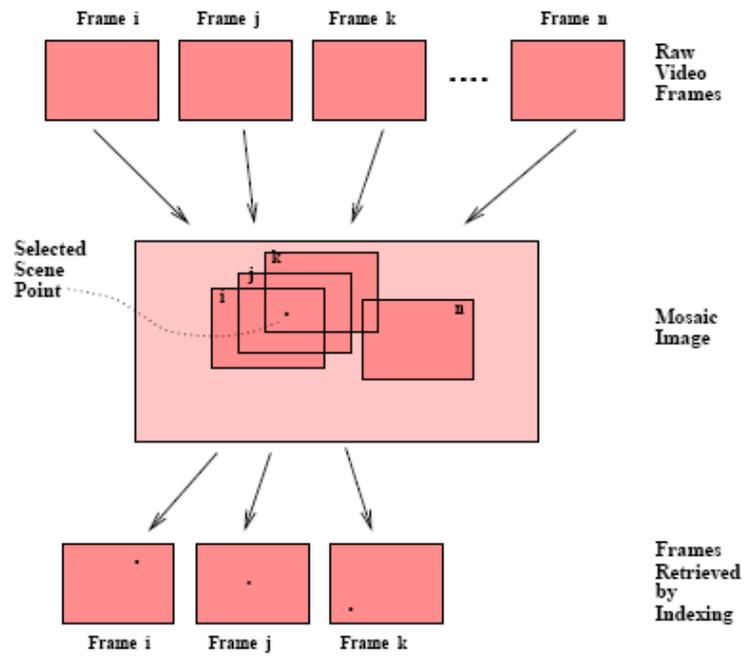


Abbildung 2.11: Indizierung: Auswahl eines Punktes Quelle:[IA98] S.9

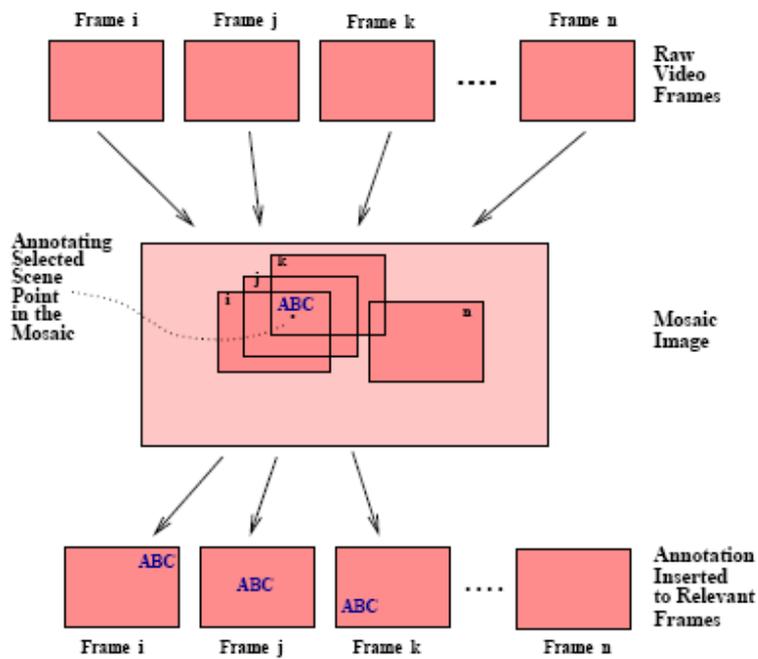


Abbildung 2.12: Vererbung der Information in alle Einzelbilder Quelle:[IA98] S.10

Inhaltsanalyse auf die Szenen-Darstellung angewendet werden. Hervorzuheben ist jedoch, dass bei der Anwendung des Video Indexing die Ergebnisse deutlich verbessert werden können. Durch die Berechnung der Bewegungsbahnen der Objekte lassen sich die Objekte im Zeitverlauf verfolgen, auch wenn diese auf einigen Einzelbildern mit klassischen Verfahren nicht zu erkennen sind.

Die Szenen-Darstellung dient als optische Inhaltsangabe zur Szene, die Summe der Szenen-Darstellungen als Inhaltsangabe des gesamten Videos.

Auf Basis der Szenen-Darstellung ist nun das effiziente Indizieren und Bearbeiten des Videos möglich. Dabei stehen zwei unterschiedlich Methoden zur Verfügung: Das ortsabhängige Indizieren und das objektbasierte Indizieren. Beim ortsabhängigen Indizieren erfolgt die Auswahl eines Punktes in der dynamischen Zusammenfassung. Über diesen Punkt können alle Einzelbilder, die diesen Punkt enthalten ausgewählt werden (siehe Abbildung 2.11). Anmerkungen zu diesem Punkt werden dann in alle Einzelbilder vererbt (siehe Abbildung 2.12). Beim objektbasierten Indizieren impliziert die Auswahl eines Objektes im Mosaik zwei Aktionen. Zum Einen, analog zum ortsabhängiges Indizieren, die Auswahl aller Einzelbilder, die das Objekt enthalten und zum Anderen die Auswahl eines Zeitpunktes, eben des Zeitpunktes an dem sich das Objekt an dem gewählten Ort befindet. Auch hierbei werden wieder die Anmerkungen zu den Objekten in diejeweiligen Einzelbilder vererbt. Somit lassen sich also eine große Zahl von Einzelbildern in einem Schritt bearbeiten.

3 Moderne Anwendungen im Entwicklungsstadium

Einen besonderen Ansatz zur Darstellung von ganzen Umgebungen verfolgt die im Entwicklungsstadium befindliche Anwendung Photosynth. Photosynth ist als Weiterentwicklung des *Photo-Tourism-Projekts*¹⁹ der Universität Washington entstanden. Inzwischen findet die Entwicklung der Anwendung im Hause des Microsoft Live Labs statt.²⁰ Es ist eine Anwendung zur Darstellung einer Bildersammlung zu einem Objekt oder einer Umgebung im dreidimensionalen Raum. Dabei werden die Einzelbilder der Sammlung auf Übereinstimmungen untersucht (u.a. unter Verwendung von SIFT²¹) und mit Hilfe der gefunden Übereinstimmungen die Anordnung der Bildebenen im Raum berechnet. Dabei wird paarweise das SIFT²²-Verfahren auf die Einzelbilder angewendet. Auch das RANSAC²³-Verfahren findet Anwendung. Über die Kombination dieser Verfahren ist die Positionsbestimmung auch umfangreicherer Bildersammlung in annehmbarer Zeit möglich (2 Beispielsammlungen: ca. 80 Fotos in 3 Stunden, ca. 600 Fotos in 2 Wochen). Über einen speziellen Navigator kann eine zusammenhängende Bildersammlung (d.h. mehrere Einzelbilder eines Motivs, bzw. einer Umgebung) betrachtet werden (siehe Abbildung 3.1). Dabei sind zahlreiche Navigationsmöglichkeiten gegeben (z.B. Zoom, Drehung, Drehung um das betrachtete Objekt, ...). Durch die Anordnung der Einzelbilder in ihrem räumlichen Zusammenhang entsteht der Eindruck der Navigation in eben dieser Umgebung.

Zurzeit stehen jedoch nur einige Beispiel-Bildersammlungen zur Verfügung.²⁴ Die Nutzung von Photosynth von Anwendern für deren eigene Bildersammlungen ist noch

¹⁹vgl. [SSS06]

²⁰<http://labs.live.com/photosynth/default.html>

²¹vgl. [Low04]

²²vgl. [Low04]

²³vgl. [FB81]

²⁴<http://labs.live.com/photosynth/collectionHome.htm>

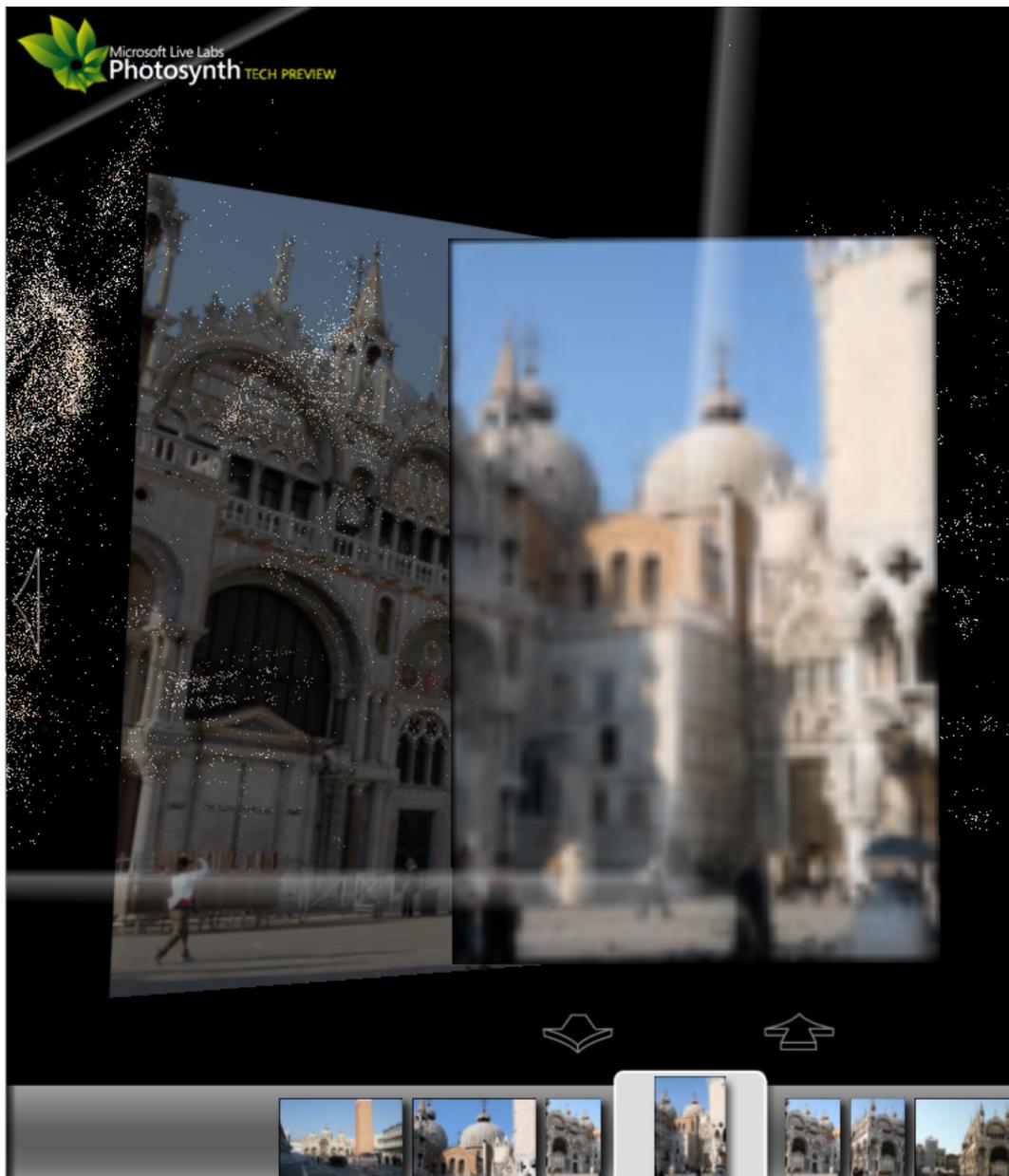


Abbildung 3.1: Photosynth

nicht möglich. Für die weitere Entwicklung der Anwendung denkt man etwa über die Verwendung von Global Positioning Data (GPS)-Daten, die heute schon in den Exif Metadaten einiger digitaler Bilddateien verfügbar ist, für die Positionsbestimmung nach. Nach Aussage von Microsoft ²⁵ ist die weitere Vermarktung und Entwicklung des Produktes noch offen. Auf Grund der stark zunehmenden Verbreitung von digitaler Fotografie im Privatbereich und damit in diesem Anwenderbereich stetig wachsenden Fotosammlungen, ist eine rege Nachfrage nach dieser Art von Anwendungen jedoch gut vorstellbar.

²⁵vgl. [Mac08]

Literaturverzeichnis

- [Che95] CHEN, Shenchang Eric. *QuickTime VR: an image-based approach to virtual environment navigation*. In: MAIR, Susan G. und COOK, Robert (Herausgeber), *SIGGRAPH '95: Proceedings of the 22nd annual conference on Computer graphics and interactive techniques*, 29–38. ACM, New York, NY, USA, 1995. ISBN 0-89791-701-4.
- [FB81] FISCHLER, Martin A. und BOLLES, Robert C. *Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography*. In: *Commun. ACM*, 24(6) 381–395, 1981. ISSN 0001-0782.
- [Gre86] GREENE, Ned. *Environment mapping and other applications of world projections*. In: *IEEE Comput. Graph. Appl.*, 6(11) 21–29, 1986. ISSN 0272-1716.
- [IA98] IRANI, M. und ANANDAN, P. *Video indexing based on mosaic representation*, 1998.
URL citeseer.ist.psu.edu/irani98video.html
- [IAB⁺96] IRANI, M.; ANANDAN, P.; BERGEN, J.; KUMAR, R. und HSU, S. *Efficient representations of video sequences and their applications*, 1996.
URL citeseer.ist.psu.edu/irani96efficient.html
- [Lip80] LIPPMAN, Andrew. *Movie-Maps: An Application of the Optical Videodisc to Computer Graphics*. In: *ACM SIGGRAPH Computer Graphics*, 14(3) Seite 32–42, 1980. ISSN 0097-8930.
- [Low04] LOWE, David G. *Distinctive Image Features from Scale-Invariant Keypoints*. In: *Int. J. Comput. Vision*, 60(2) 91–110, 2004. ISSN 0920-5691.

- [Mac08] MACINTYRE, Jeffrey. *Starren und staunen*. In: *Technology Review - Heise Verlag*, (4/2008), 2008.
- [Rip89] RIPLEY, G. David. *DVI a digital multimedia technology*. In: *Commun. ACM*, 32(7) 811–822, 1989. ISSN 0001-0782.
- [SSS06] SNAVELY, Noah; SEITZ, Steven M. und SZELISKI, Richard. *Photo tourism: exploring photo collections in 3D*. In: *SIGGRAPH '06: ACM SIGGRAPH 2006 Papers*, Seite 835–846. ACM, Conference Chair: John Finnegan and Program Chair: Julie Dorsey, New York, NY, USA, 2006. ISBN 1-59593-364-6.
- [WWDM99] WRIGHT, Mark; WATSON, Gordon; DUNLOP, Gair und MIDDLETON, Roy. *Edinburgh: 200 years of heritage through image-based virtual environments*. In: *Proceedings of 5th INTERNATIONAL CONFERENCE ON VIRTUAL SYSTEMS AND MULTIMEDIA*. VSMM Society, 99.