

注釈情報付き全周映像システムの開発

山本 強[†] 柴田 幸祐[†] 茂木 達也[†]

[†]北海道大学大学院情報科学研究科

〒060-0814 札幌市北区北14条西9丁目

E-mail: [†]yamamoto@ist.hokudai.ac.jp, {siba, mogi}@nis-ei.eng.hokudai.ac.jp

あらまし 超高速ネットワーク環境の普及により、コンテンツの高品質化、大容量化だけでなく、これまで実現できなかった新しいサービスが可能になっている。高速ネットワーク環境の普及を想定し、我々はネットワーク上でリアルタイム中継可能な全周映像システム PanoVi を開発している。PanoVi は一般の広角カメラ4台で360° 全周の動画像を収録し、1映像ストリームに画面合成して伝送し、再生側で3次元CGとして任意方向の映像を再構成する。本報告では、PanoVi の基本モデルと、それに対して座標やコメントなどの注釈情報を付加することによって実現される新しい情報表現モデルと、その実装を報告する。

キーワード パノラマ映像, XML, GIS, Direct-X, テクスチャマッピング, 注釈, メタデータ

Development of Panoramic Movie System with Annotation

Tsuyoshi YAMAMOTO[†] Kohsuke SHIBATA[†] and Tatsuya MOGI[†]

[†] Graduate School of Information Science and Technology, Hokkaido University

N-14 W-9, Kita-ku, Sapporo, 060-0814 Japan

yamamoto@ist.hokudai.ac.jp, {siba, mogi}@nis-ei.eng.hokudai.ac.jp

Abstract As the bandwidth of the network increase, it comes to reality to create a new video presentation model that we are not familiar until now. In this report, we introduce PanoVi, a new panorama video capturing and delivering system that is applicable to up-to-date network environment. PanoVi captures movie with 360 degree of viewing angle by using 4 commodity TV cameras and compose into a single standard video stream and reconstruct movie of arbitrary viewing direction by using 3D graphics accelerator hardware. In this report, we describe system architecture and implementation of PanoVi. Further, we introduce new application that is realized by using annotation or meta-data contents.

Keyword Panorama Movie, XML, GIS, Direct-X, Texture Mapping, Annotation, Meta-data

1. まえがき

近年臨場感通信、仮想空間の構築、ロボットの遠隔操作あるいは防犯などの需要から、広い視野角を有した映像の収録・再生システムについての研究が盛んに行われている。IBR(Image based Rendering)の手法による視点自由度の高い画像表現技術が実用化されたため、現在では静止画に関しては視点自由度が2のパノラマ画像から5のPlenoptic Function[1]までがシステムとして提案されている。全周画像フォーマットで最も成功したのは QuickTime-VR[2]に代表されるモザイク合成全周画像方式であろう。この方式はある点から撮影された視軸方向の異なる画像を連結し、円筒座標や極座標系でシームレスな画像として表現し、再生時に

切り出しと画像歪を補正するものである。

全周画像技術の自然な拡張として動画像ベースのパノラマ映像方式の実現が望まれている。これまでもテーマパーク等では全周映像が使われているが、多くの場合は特殊光学系を用いて収録されたフィルム映像として作成されたものである場合が普通であり、ネットワーク経由で高品質にリアルタイム伝送可能な全周映像システムはこれまでほとんど実現されていなかった。

我々はこれまで4台のカメラで全周映像を収録し、リアルタイム中継・再生可能な全周動画像システム PanoVi[3,4]を開発し・公開している。PanoViは複数カメラで全周を一部重複した映像群として撮影し、1ストリームの映像フォーマットに合成した後で記録・伝送し、再生側で光学系の幾何学的ゆがみ、接合部にお

ける輝度・色相の不連続をテクスチャマッピングを用いて再生時にデジタル補正して違和感のない全周映像を実現する。

2. スケーラブル構成の全周動画像システム -PanoVi-

単一カメラによる全周動画像収録は再生解像度が著しく低いため、全周映像が目的とする臨場感が満足に得られるものにはならない。よりリアリティのある全周映像を実現するためには、ハイビジョンを上回る高解像度の画像記録方式が必要となる。自然な解として、ハイビジョンカメラと魚眼レンズ等の超広角光学系を用いる方式が考えられるが、単一光学系で180度以上のパノラマを収録することは事実上不可能である。現実解としては複数カメラの映像を記録座標系(極座標、円筒座標)でモザイク合成し、高い空間解像度を実現する方法がある。

複数カメラを用いる方式は、6面体スクリーンを用いるCAVEでの全周映像記録再生方式として用いられている。これは、CAVEの壁面と同一方向の最大6台のカメラを用いて全周を撮影し、それをCAVEの壁面に再投影する極めて単純な仕組みである。光学系が精密に作られていれば座標空間での処理は不用であるものの、CAVEは再生環境が

特殊であり、一般向けコンテンツの再生環境とは言いがたい。

複数カメラを用いて撮影された多数の画像をより広い画像空間でシームレスに接続し等価的に超広角画像を合成する手法は一般にイメージモザイクと呼ばれ、パノラマ写真の合成に実用化されている。この方法は低解像度のカメラを用いても高品質の超広角画像が記録できる可能性がある。しかし、動画像に対するモザイク合成処理は莫大な計算量が必要とするため、実現が困難であった。

我々は動画像で多眼モザイク合成を行う全周映像収録・再生システム PanoVi を開発している。PanoVi は超小型の広角ビデオカメラ(画角 114°)を4台用いて全方向の動画像を収録し、再生側でモザイク合成を行いシームレスな全周映像を合成するシステムである。PanoVi はモザイク合成を再生側で行うことによって収録システムが単純な構造で実現できるようになった。その結果屋外での収録や移動体での収録も可能になっている。また、中間映像フォーマット(伝送フォーマット)をNTSCやHDTVといった標準フォーマットに設定することができるため、既存の映像中継システムを用いて実時間中継できるように構成している。図1にPanoViのシステム概念図を示す。

全周動画像でのモザイク合成はレンズの光学歪の補正と接続部の位置あわせ処理の2段階で行われる。これまで、モザイク合成処理は2次元画像処理として

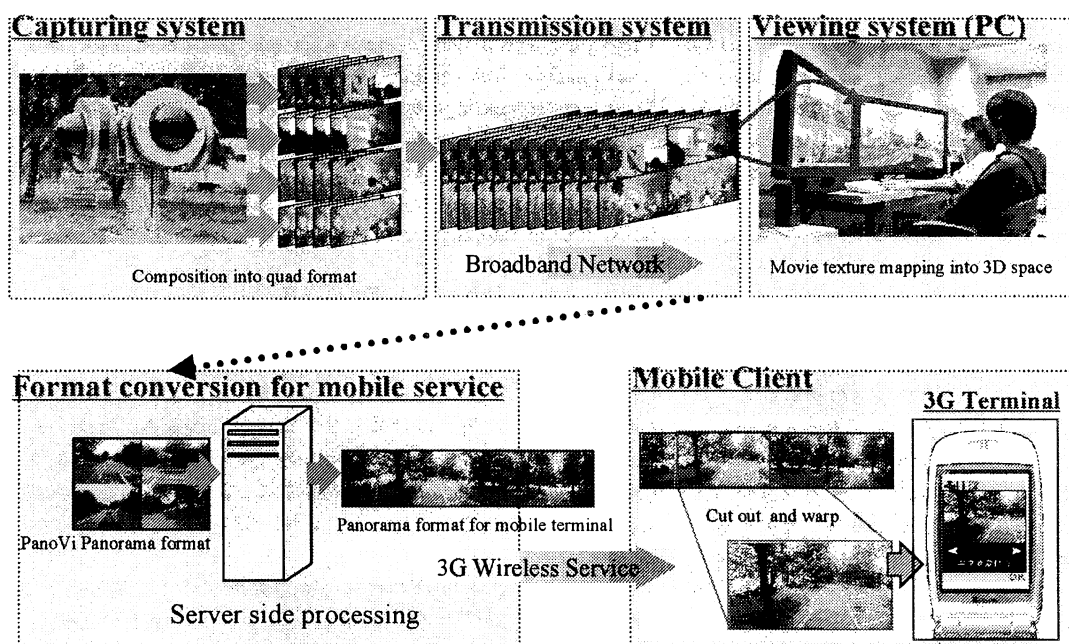


図1 全周動画像システム PanoVi のアーキテクチャ

実装されることが普通であったが、PanoVi では 3 次元空間での動画プロジェティブテクスチャとして合成処理を行う設計になっている。このため、リアルタイム動作には高レベルの 3D グラフィクスアクセラレータが必須となるが、最近では低価格デスクトップ PC やノート PC でも 3 DCG API に対応した高速アクセラレータを装備しており、ハードウェアの制約は少なくなっている。

図 2 に 4 台のカメラで収録した全周映像から任意方向に視軸を設定した仮想カメラ映像を合成する処理モデルを示す。4 台のカメラ映像は単純にタイル合成された中間フォーマットを伝送形式としてネットワークやファイル形態で再生クライアントに送られる。再生クライアントでは、3 次元空間に全周スクリーンを微小な四角形ポリゴンで記述し、各微小ポリゴンに対する動画テクスチャを 4 画面合成映像内の任意の 4 角形エリアに対応させる処理を行う。仮想スクリーンはカメラ台数に分割され、隣接する仮想スクリーンと接合部分はオーバーラップするように配置され、オーバーラップ部分はアルファブレンド処理により隣接カメラ映像とスムーズに接続する。これにより、多カメラ撮影の全周映像での接続部の不連続を緩和し、違和感の少ない全周映像を再構成することができる。

これら一連の処理は Windows における 3 次元グラフィックス API である Direct-X 9 で標準 API として準備されており一般的なグラフィックスハードウェアで、NTSC 解像度の動画テクスチャを毎秒 30 フレームでマッピング可能ことが確認されている。システム組み込みの仮想スクリーン形状としては 4 平面モデルと円筒モデルが用意しているが、全周スクリーンの形状は任意に構成可能であり、撮影座標系が変わった場合にも、仮想スクリーン形状を再設計することで違和感のない多視点映像環境を合成することができる。

パノラマ動画では毎秒 30 フレームで複数画像の幾何学変換と合成を行う必要があり、これまで汎用 PC レベルの処理速度でリアルタイム処理は困難であったが、近年 PC に組み込まれるグラフィクスアクセラレータの高度化、3DCG 用 API の高レベル化によりリアルタイム処理が可能になっている。また、CPU の高速化により MPEG 再生処理と他のアプリケーションを並行して実行することも可能となり、全周動画画像を含んだアプリケーションシステムの実現も可能となってきた。

PanoVi の設計上の特徴は、それを構成するサブシステムが既存デバイスの組み合わせで実現されていることである。具体的には、撮影カメラは一般的な広角 TV

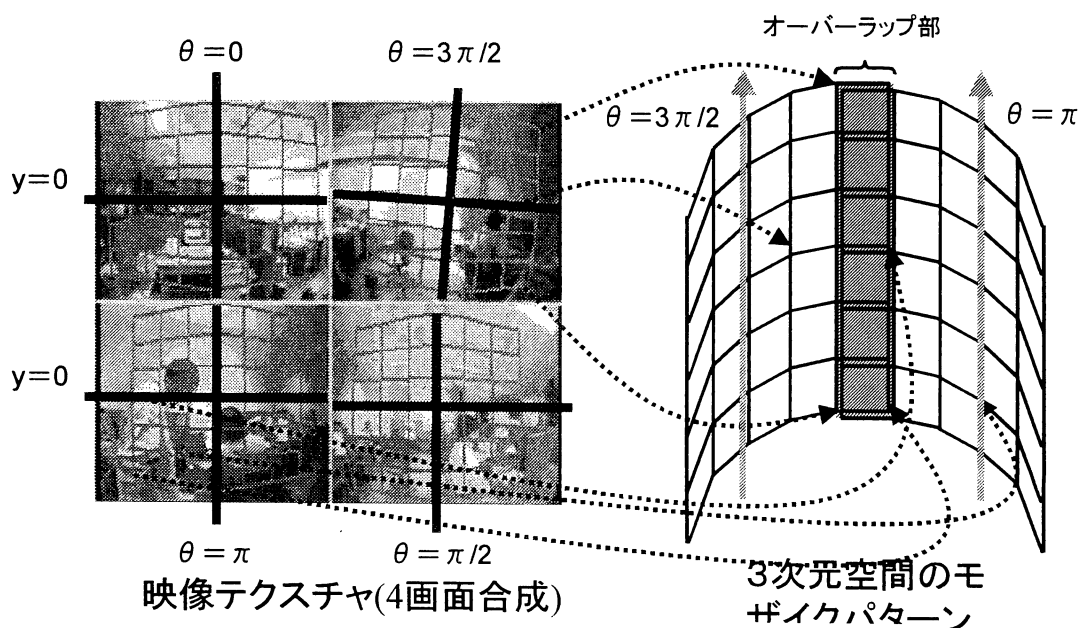


図 2 テクスチャマッピングによる光学歪補正及び接続

```

<!DOCTYPE PANОВI_MAP [
  <!-- ルート情報 -->
  <!ELEMENT PANОВI_MAP (MAPINFO)+>
  <!-- 注釈基本情報の構造 -->
  <!ELEMENT MAPINFO (MOVIE, AREA, DATA) >
  <!-- ムービー情報 -->
  <!ELEMENT MOVIE(FILENAME, ENCODING,
DURATION) >
  <!ELEMENT FILENAME (#PCDATA)>
  <!ELEMENT ENCODING (#PCDATA)>
  <!ELEMENT DURATION (#PCDATA)>
  <!ATTLIST DURATION UNIT #PCDATA>
  <!-- エリア情報 -->
  <!ELEMENT AREA (LT, RB) >
  <!ATTLIST AREA TYPE #PCDATA >
  <!ELEMENT LT (NL, EL) >
  <!ELEMENT RB (NL, EL) >
  <!-- 注釈データ -->
  <!ELEMENT DATA (ROUTE)+ >
  <!ELEMENT ROUTE (POINT)+>
  <!ATTLIST ROUTE NOTE #PCDATA>
  <!ELEMENT POINT (NL, EL, TIME) >
  <!-- 座標、時間 -->
  <!ELEMENT NL (#PCDATA)>
  <!ELEMENT EL (#PCDATA)>
  <!ELEMENT TIME (#PCDATA) >
  <!ATTLIST TIME UNIT #PCDATA>
] >

```

図 3 座標注釈用 XML の DTD 定義

カメラを使用し、記録・伝送フォーマットも NTSC や MPEG を想定している。再生環境は汎用 PC+グラフィックスアクセラレータと汎用グラフィックスライブラリのみで実現」されている。そのため、技術革新によって構成要素の性能指数が向上した場合に速やかにその効果を取り込むことが可能となっている。

3. 全周映像と注釈情報

画像、映像は主観的な情報コンテンツとして魅力的であり、信号としての映像品質や表現力に注目が集まっている。しかし、ネットワーク上で統合された情報サービスの構成要素として新しい映像システムを考えるなら、検索システムへの対応、他の情報コンテンツとの関連、映像に対する注釈情報が必要になる。関連する動きとして、TV 番組表の電子化 (EPG) や映像コンテンツに対するメタデータ体系の策定の動きもあり、商用コンテンツの分野で今後標準化が進むと予想され

```

<?xml version="1.0" ?>
<panovi_map version="0.1">
  <movie>
    <filename>hokudai_hi_1.mpg</filename>
    <encoding>MPEG1</encoding>
    <duration unit="sec">3310</duration>
  </movie>
  <area type="rectangle">
    <lt>
      <nl>43.04.21.70</nl>
      <el>141.20.23.10</el>
    </lt>
    <rb>
      <nl></nl>
      <el></el>
    </rb>
  </area>
  <data>
    <route note="北大工学部教養～工学部">
      <point note="北大工学部裏">
        <el>141.20.35.232</el>
        <nl>43.04.28.503</nl>
        <time unit="msec">15000</time>
      </point>
      <point note="北大工学部裏・化学棟付近">
        <el>141.20.35.712</el>
        <nl>43.04.28.795</nl>
        <time unit="msec">19000</time>
      </point>
      <point note="北大工学部裏・化学棟東口">
        <el>141.20.35.912</el>
        <nl>43.04.29.087</nl>
        <time unit="msec">24000</time>
      </point>
    </route>
  </data>
</panovi_map>

```

図 4 北大キャンパスの撮影済み全周コンテンツに対する座標注釈 XML の例

ている。

一般的な意味でのメタデータは、コンテンツパッケージ全体に対する注釈、つまり大粒度の注釈であると考えられる。これに対して、チャプタータイトルや吹き替え字幕などは粒度が中レベルから小レベルの注釈であるといえる。映像に対する注釈付加単位は時間(フレーム)、画像座標、画像中のオブジェクトなど各種レ

ベルがあり、今後統一した記述モデルが必要となるが、それは今後の課題として、我々が対象としている全周動画像という映像コンテンツに対する注釈としていくつかの具体例を紹介し、それが可能にする各種のサービスを紹介する。

3.1 全周動画像に対する座標注釈とその応用

全周動画像システムは再生クライアント側で視点移動の自由度が2（視軸方向 θ 、 ϕ ）の映像システムである。

パノラマ映像のように視点自由度が高い映像システムは、自由度が無い通常の映像よりもより高度であると思われがちだが、自由度が高いために何らかの補助情報を与えなければ、そのコンテンツの意味が理解できないという状況が出てくる可能性がある。高自由度のコンテンツこそその内容を記述するメタデータの有無がその利用価値を決めると言える。

全周動画像の応用として、定速度で移動しながら道路沿いに撮影し、コンテンツとする応用モデルがある。この場合、映像フレームと撮影中心位置の対応情報を全周動画像に対するメタデータとすることが考えられる。単純にはフレームごとに撮影位置とカメラ基準方

向の情報を付加することになるが、車載カメラで一定速度で移動しながら撮影する場合は等速移動区間については区間の始点と終点についてフレーム番号と座標のペアを定義することにより、区間内は線形補間により位置を推定することができる。このモデルによれば、30分程度の全周映像コンテンツに対する座標注釈情報はXMLで数KBに収めることができる。

図3、4は、我々が例題として定義した座標注釈情報のDTD定義と記述サンプルである。これだけのメタデータを付加することで、移動撮影した全周映像コンテンツを地図と関連づけることができ、たとえば地図にある道路の任意の位置を指示することにより、その位置から撮影された全周画像フレームを切り出すことができる。

我々は北海道大学キャンパス内の主要道路の全周映像を撮影し、それに対する座標メタデータを付加すると同時に、そのスキーマ及びアプリケーションを作成した。

図5は我々が試作したアプリケーションシステムの全体構成である。座標注釈データは撮影時にGPSなどで同時収録が可能であるが、この実験では、全周映像コンテンツのパッケージ化後にマニュアルでメタ

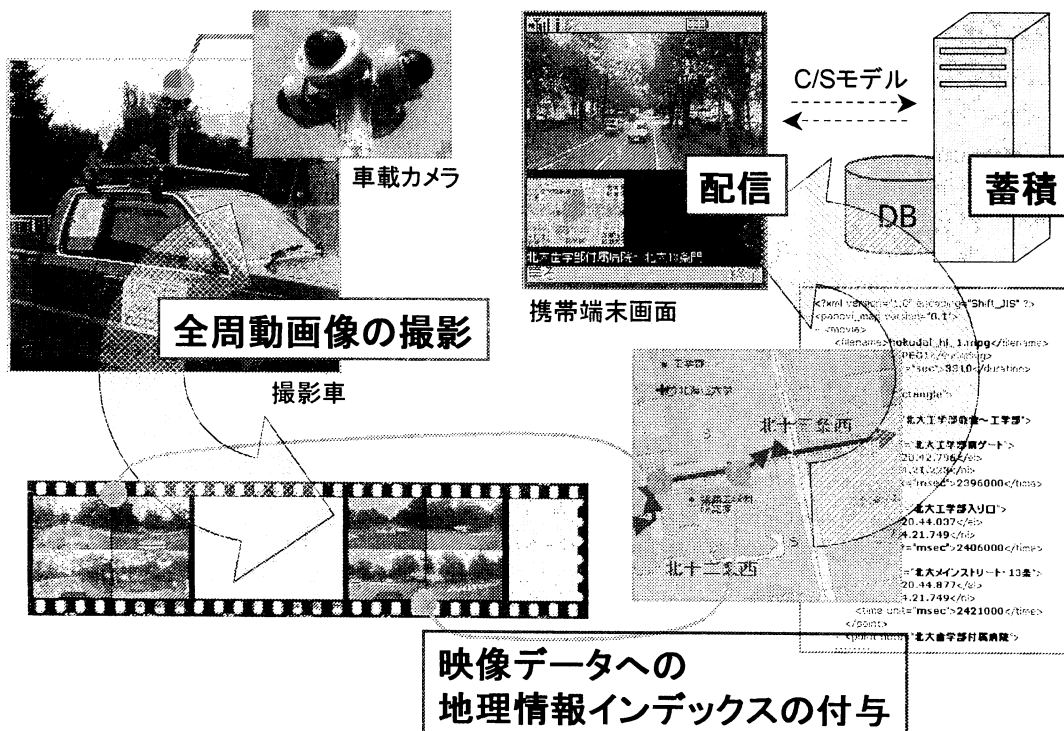


図5. 座標注釈データを応用した地図-全周動画像の連携サービス

ータを記述している。そのため、タイムフレームと座標の対応には誤差が発生しているが、アプリケーションでは幾何ひずみがあるイラスト地図をベース地図として用いることを想定しているため、座標誤差はあまり問題となっていない。

全周映像は大容量コンテンツであり、高速ネットワークと高性能 PC を対象として開発されたが、座標注釈を付加し、地図情報との関連を定義することで静止画像の切り出しが容易になり、動画として収録されたコンテンツを携帯電話など狭帯域回線で接続される端末向けサービスとして転用する可能性を示すことができた。

4. むすび

我々が開発している全周動画システム PanoVi のコンテンツに注釈データとして撮影座標情報を付加する方式と、それを応用するアプリケーションの実装例を紹介した。PanoVi は再生クライアント側で 3 次元空間への画像マッピングにより画像再構成を行っているため、座標と関連した各種注釈情報を 3 次元空間に配置し、画像合成時に反映させることが可能である。これまで 2 次元の画像処理として実現されていた全周画像・映像と比較して PanoVi は空間記述レベルが 1 段高いという見方もでき、今後注釈情報を再生画面に反映するモデルや実装について開発を進める計画である。

文 献

- [1] L. McMillan, G. Bishop "Plenoptic Modeling: An image-based rendering system," *Proceedings of SIGGRAPH'95* pp.36-46 (1995)
- [2] S.E. Chen, "QuickTimeVR - An Image-Based Approach to Virtual Environment Navigation," *Proceedings of SIGGRAPH'95*, pp.29-38(1995)
- [3] R. Szeliski, H.Y. Shum, "Creating full-view panoramic image mosaics and environment map," *Proceedings of SIGGRAPH'97*, pp.251-258(1997)
- [4] M. Doi and T. Yamamoto: " PanoVi: A Multi-Camera Panoramic Movie System by using Client-Side Image Mosaicking , " *Proceedings of IASTED Modelling and Simulation MS'2003*, 380-225, (2003)
- [5] T. Yamamoto and M Doi: Design and Implementation of Panoramic Movie System by Using Commodity 3D Graphics Hardware, *Proceedings of CGI2003*, Jul. 2003 Signal Processing, pp. 389-394(2003)