

光線空間法と3次元映像取得・表示方式

藤井俊彰^{†1}

光線空間法とは、3次元空間の視覚的な情報を、空間を伝播する光線の情報として表現する手法であり、さまざまな3次元映像入出力方式を統一的に表現する理論的基礎として有用であるとともに、ライト・フィールド・レンダリング (LFR) と称されるフォトリアリスティックな映像生成や自由視点テレビ (FTV) における中核技術として応用されている。本稿では、光線空間法概念・定義から始め、現在話題となっている3次元映像や多視点映像との関係について解説するとともに、3次元・多視点映像の取得・表示技術の現状と展望を述べる。さらに、光線空間法の課題である補間と情報圧縮について、近年の研究や国際標準化の動きについて述べる。

Ray-Space Theory and It's Application to 3D Video Acquisition and Display

TOSHIAKI FUJII ^{†1}

This paper presents the "Ray-Space" theory and its applications to photo-realistic image synthesis used in Light Field Rendering and Freeviewpoint TV. Ray-Space provides a common representation for various types of 3D images and can be applied to a common data format for the Integrated 3-D Visual Communications. This paper starts with the concept and definition of Ray-Space and describes relationship between Ray-Space and 3D/multiview videos. Acquisition and display systems of 3D/multiview video are then reviewed in terms of Ray-Space sampling and display. We also describe Ray-Space interpolation and compression that are key technologies to realize ultimate 3D/multiview communications. International standardization activities (MPEG) for multiview video coding are also presented.

^{†1} 東京工業大学 大学院理工学研究科 集積システム専攻
Tokyo Institute of Technology

1. はじめに

2010年は「3D元年」と言われ、各社が相次いで3D対応テレビを発表しており、次は眼鏡なしの3Dへ、さらには多視点や自由視点映像の実現へと期待が高まっている。また、近年のPCやデジタルカメラなどの画像取得装置の高性能化・低価格化に伴って3次元映像・多視点映像の取得装置が安価に構築できる環境が整ってきたことから、3次元映像処理・多視点画像処理に関する研究も多く見られるようになってきた^{1),2)}。また映像・音声の圧縮方式の国際標準化を定める会合であるMPEG(Moving Picture Experts Group)においても多視点映像符号化方式の国際標準が策定され、さらに今後に向けて3次元ビデオや自由視点テレビ (FTV)^{3),4)}などの標準化が活発に議論されている⁵⁾。

このような3次元映像・自由視点映像の入出力方式を統一的に表現する手法として、光線空間法がある。これは、3次元空間の視覚的な情報を、空間を伝播する光線の情報として表現する手法であり、さまざまな3次元映像を統一的に取り扱う理論的基礎として導入された。その後、フォトリアリスティックな映像生成や自由視点テレビにおける映像生成に応用されている。

本稿では、まず3次元映像処理の基礎となる概念である光線空間の概要を紹介し、その応用としての自由視点映像の生成、可変焦点画像の生成について述べる。また、近年のさまざまな3次元映像の取得・表示手法を光線空間の取得・表示の観点から述べる。最後に、光線空間の処理における重要な課題である補間と情報圧縮について最近の研究、今後の展望について述べる。

2. 光線空間の導入

2.1 3次元統合画像通信

3次元映像表示技術にはさまざまなものがある。現在、映画やテレビ・DVD等の市場で活況を呈しているのは2眼のメガネ式である。高解像度・フルカラーの3次元映像の表示が可能なものとして普及が期待されている。また、レンチキュラーレンズやパララックスバリアなどのメガネなしの3次元映像方式も徐々に市場に出まわりつつあり、携帯電話やゲーム機などへの応用が試みられている。一方、撮影システムとしては、2眼のカメラシステムや多視点撮影デバイスなどが市場でも見られるようになってきた。コンテンツとしては、現時点ではコンピュータグラフィックスで作成したものが多いが、撮影システムの発展に伴って実写のコンテンツも充実してきつつある。このように撮影システムから表示システムま

でさまざまなシステムが乱立しつつあり、それらの相互接続性を保証することが困難になっている。またコンテンツ生成者においても、どのようなフォーマットで作成すればよいかを課題となっている。

さて、このような状況下での3次元映像の通信システムを考える場合、どのような方式を採用すればよいであろうか。現時点では、2眼のシステムにおいて従来の2次元画像通信システムとの互換性を念頭に、2眼映像を左右に並べたサイド・バイ・サイド方式や上下に並べたトップ・アンド・ボトム方式が一部で採用されつつある。しかしながら将来に目を向けた場合、2眼式のみならず、多眼式のシステムについても考慮する必要が出てくるであろう。これらは将来的に一つの方式に集約されていくというよりは、各々の特徴を生かせる場所所で用いられていくことになるであろう。

原島らは1990年代初頭にこのような将来像を予測し、従来のデバイス中心の通信方式ではなく、さまざまな方式の3次元映像方式が共存することを前提とした「3次元統合画像通信^{6),9)}」(Integrated 3-D Visual Communication)を提案した。これは3次元映像の中間的なフォーマットを定め、入出力デバイスを独立に規定しようという試みである。図1に3次元統合画像通信の概念図を示す。この構想では、2眼式、多眼式、CGデータ、さらにはホログラフィーに至るまでのさまざまな3次元映像方式に対する中間的なフォーマットを規定する。入力側ではそれらを中間フォーマットに変換・符号化した上で伝送・蓄積を行い、出

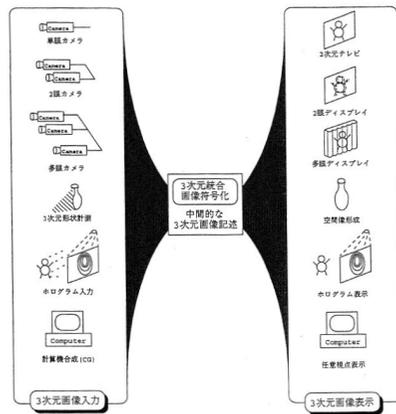


図1 3次元統合画像通信⁹⁾

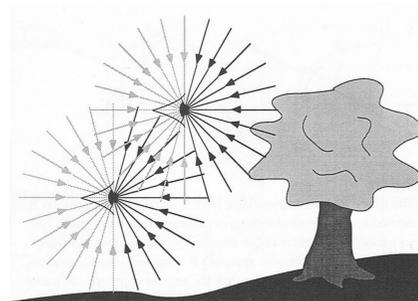


図2 Plenoptic Function⁷⁾

力側では中間フォーマットを元に各種の3次元ディスプレイ方式で規定されたフォーマットへと変換して出力する。この中間的なフォーマットとしてどのようなものがふさわしいかという問題を検討し、得られた一つの解答が以下に述べる光線空間の概念である。

2.2 光線空間の定義

光線空間を導入するに当たり、まず Adelson らによって提案された Plenoptic Function について述べておく⁷⁾。人間が図2のようなシーンを観測するとき、人間の初期視覚では、以下のような関数で表される量が観測されるとした。

$$P = P(x, y, \lambda, t, V_x, V_y, V_z) \quad (1)$$

(1)式が意味するところは、視点位置 (V_x, V_y, V_z) に目を置いたとき時間 t において画像座標 (x, y) の位置に波長 λ の光が P の強度で観測される、ということである。そして人間の初期視覚とは、この7次元空間で表される情報を取得し、エッジ、コーナー、色、動きなどの検出結果の組み合わせを出力するものであると定式化した。

Plenoptic Function は確かに視覚への入力 of 全て (plenus) を表してはいるが、3次元画像の入力・出力を考える際にここまで自由度を高くしても有用な結果は得られない。そこで、光の反射・屈折を考えない自由空間においては光は減衰せず直進するものと仮定し、幾何光学的な意味での「光線」を考える。この概念においては、3次元シーンは光線で満ちており人間の目はそれを観測する観測器であると考えられる。前述のように様々な形態の3次元映像入出力方式に対応した中立的な記述法とは、3次元シーンに満ちている光線に着目してこれを記述するものとし、3次元空間情報を光線のパラメータ空間上で定義された関数で表現することを提案した⁸⁾。この光線パラメータ空間を「光線空間 (Ray-Space)」と呼ぶ⁹⁾⁻¹²⁾。

さて、図3に従って光線空間を定義する。図中の3次元実空間 $O-X-Y-Z$ において基準平面 $Z=0$ を考え、空間を伝播する「光線」の基準平面を通過する位置 (x, y) と光線の通過方向 (θ, ϕ) で表す。これら4つのパラメータにより、実空間中を伝播する光線を一意に表すことができる。この光線の光強度 (観測された輝度値またはRGB値) を f とすると、光線の表現として4次元空間で定義された関数

$$f(x, y, \theta, \phi) \quad (2)$$

を考えることができる。これは3次元空間中に満ちている光線をパラメータ表現し、そのパラメータに対して輝度値を出力する関数として定義したものである。人間の目は光線観測装置と考えることができるため、この関数表現により、物体の形状や位置・奥行き、さらには鏡面反射などの表面特性などを記述することなく、3次元シーンの「見え」の情報を全て記述することができる。

上記の表現方法では、基準面 ($Z = 0$) より観測者側に射出してくる光線のみを記述しているため、基準面と平行な、もしくは反対側から見た情報を記述することができず、物体を全周囲から観察したシーンの記述などの場合に不便である。そこで、光線を直交座標ではなく極座標で定義した表記も与えておく¹³⁾。光線の方法は(2)式と同じ (θ, ϕ) で表し、着目している光線と基準面が常に垂直になるように $Z = 0$ 平面を (θ, ϕ) だけ回転させる。元の座標系 (X, Y) を回転させた座標系を (ξ, η) とすると、この光線空間は以下のように表される。

$$f(\xi, \eta, \theta, \phi) \quad (3)$$

(2)式で定義されたものを直交座標光線空間 (Orthogonal Ray-Space), (3)式で定義されたものを極座標光線空間 (Spherical Ray-Space) と呼ぶ。

2.3 実空間と光線空間の関係

次に、実空間と光線空間の関係について考える。簡単のため、縦方向の視差を考えない(光線の仰角/俯角を無視する; $\phi = 0$) とし、図4に示すように直交光線空間について $Y = \text{const.}$ における実空間と光線空間の断面を考える。さて、いま実空間中の一点 $P(X, Z)$ を通過する光線を考え、これを光線空間にマッピングする。例えば、図4において光線 a の光線パラメータは、光線の X 軸との交点の座標 x_a と Z 軸正方向となす角度 θ_a の二つである。これを光線空間 (x, θ) にマッピングし、 (x_a, θ_a) における関数値 f としてこの光線の輝度値を与えるものとする。光線 b, c も同様にして光線空間にマッピングしていく。このように、点 $P(X, Z)$ を通過する「光線群」を考え、それらの光線空間 $f(x, \theta)$ における軌跡を導出

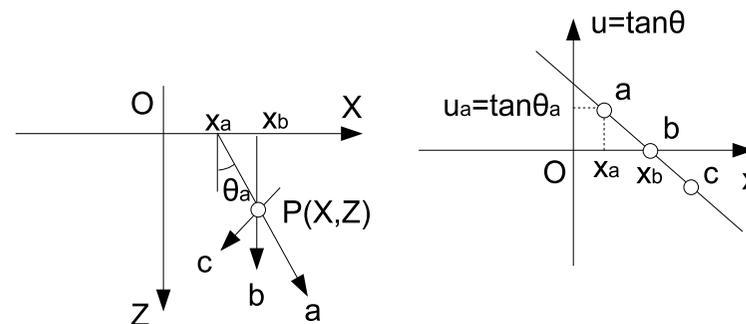


図4 実空間と(直交座標)光線空間の関係(左:実空間, 右:光線空間)

すると、幾何的な考察により

$$X = x + Zu, \quad u = \tan \theta. \quad (4)$$

が得られる。

すなわち、実空間内の一点 (X, Z) を通過する光線群は、光線空間 $f(x, u)$ ($u = \tan \theta$) 上で直線を描くことが分かる。

一方、極座標光線空間においては $\phi = 0, Y = \text{const.}$ における実空間と光線空間の関係は次のようになる。実空間中の一点を $P(X, Z) = P(r \sin \alpha, r \cos \alpha)$ と極座標表示し、この点を通過する光線群を考える。それらが光線空間 $f(\xi, \theta)$ において描く軌跡を導出すると、

$$\xi = -r \sin(\theta - \alpha) \quad (5)$$

となる¹³⁾。

さて、このように得られた実空間と光線空間との関係は、一般的に縦方向の視差まで考えた $f(x, y, \theta, \phi)$, $f(\xi, \eta, \theta, \phi)$ に容易に拡張することができる。すなわち、実空間中の一点を $P(X, Y, Z)$ とし、この点を通る光線群が4次元光線空間で描く軌跡を考える。光線は $P(X, Y, Z)$ を通るという制約があるため、自由度としては光線方向に相当する (θ, ϕ) のみである。これは (θ, ϕ) を与えることにより、 (x, y) もしくは (ξ, η) が決まることを意味する。(4)式や(5)式は、この関係の低次元版である。これらをまとめると、以下の結論が得られる。

3次元実空間中の一点を通る光線群の軌跡は、4次元光線空間のうちの2次元部分空間を構成する。

このように空間中の一点を通る光線群を考える意義は、以下のような状況を考える際に現

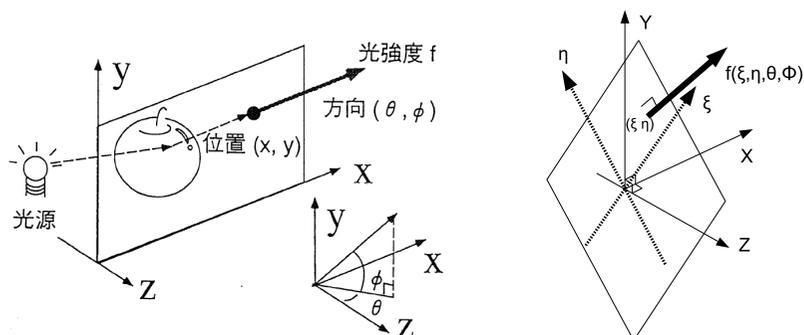


図3 光線空間の定義(左:直交座標⁹⁾⁻¹²⁾, 右:極座標¹³⁾)

れる。

(1) 物体表面の一点から射出される光線群

(2) ピンホールカメラモデルにおけるピンホールに入射する光線群

(1) は、物体表面が拡散反射面 (Lambert 面) であった場合、そこから射出される光線群の輝度値は非常に近くなるため、後述するように光線空間データが直線・sin カーブといった特徴的な構造を持つことに反映される。(2) は、シーン中にピンホールカメラを置いた場合に撮影される画像を構成する光線群である。すなわち、一点を通る光線群を考えることにより、4次元光線空間は

- 輝度値の近いデータが2次元部分空間内に並ぶ構造を持ち、
- 2次元部分空間を取り出すことにより画像として観測できる

という特徴を持つことが分かる。

図5に $f(x, u)$ および $f(\xi, \theta)$ の断面図の例を示す。直交光線空間では物体の1点から発せられた光は光線空間内で直線上に並ぶが、物体表面は Lambert の反射則に従うことが多く、従って直線上にほぼ同じ輝度値のデータが並ぶこととなる。さらに、観測者から遠い位置にある物体が手前の物体によって隠されるオクルージョンが生じることがある。従って、図5のようにさまざまな傾きの直線からなり、かつ手前の物体が奥の物体を隠すような構造が観測される。一方、極座標光線空間においては軌跡は周期が 2π の sin カーブとなり、物体の点の位置に依存して振幅と位相が異なるさまざまな sin カーブの重なりを持つている。

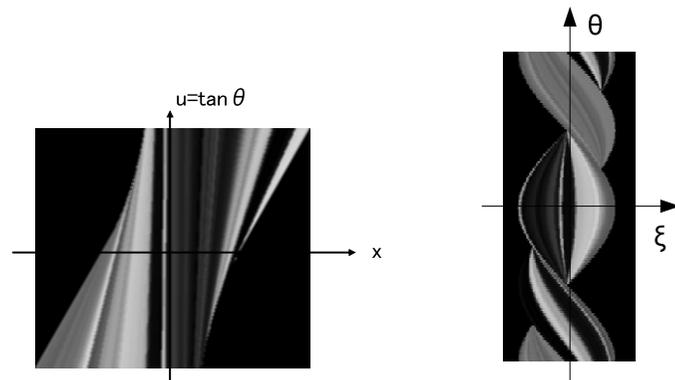


図5 光線空間の断面図の例 (左:直交座標, 右:極座標)

図5に示すような光線空間の断面図は、Bolles らにより提案された EPI(Epipolar Plane Image)¹⁴⁾ と本質的に同じものである。ただし、Bolles らの EPI は Structure from Motion の枠組みで提案されたものであり、カメラをシーン中で動かした際に観測されるエピポーラ線上の画素値の並びをカメラ位置の関数としてプロットしたものである。従って、必然的に EPI のパターンはカメラの動きに依存したものとなる。一方、光線空間とはシーンのワールド座標系で定義された光線のパラメータが張る空間のことであるから、カメラの動きや、どのように観測したか、という要因とは独立に定義されるものである。すなわち、観測するカメラの外部パラメータや内部パラメータに依存することなく、シーンそのものに付与された構造であることに注意が必要である。

2.4 他の類似の手法との関係

コンピュータグラフィックスの分野で、フォトリアリスティックな映像生成法として Light Field Rendering¹⁵⁾ が1996年の SIGGRAPH で発表され、注目を浴びた。光線空間が基準面を通る位置と方向で光線を記述したのに対し、Light Field では2枚の基準面 s - t 平面、 u - v 平面を用いている。両者は光線パラメータを定義する座標系の違いを除いて同一のものであり相互に変換が可能である。特に、Light Field において、光線角度を一定に保ちつつ

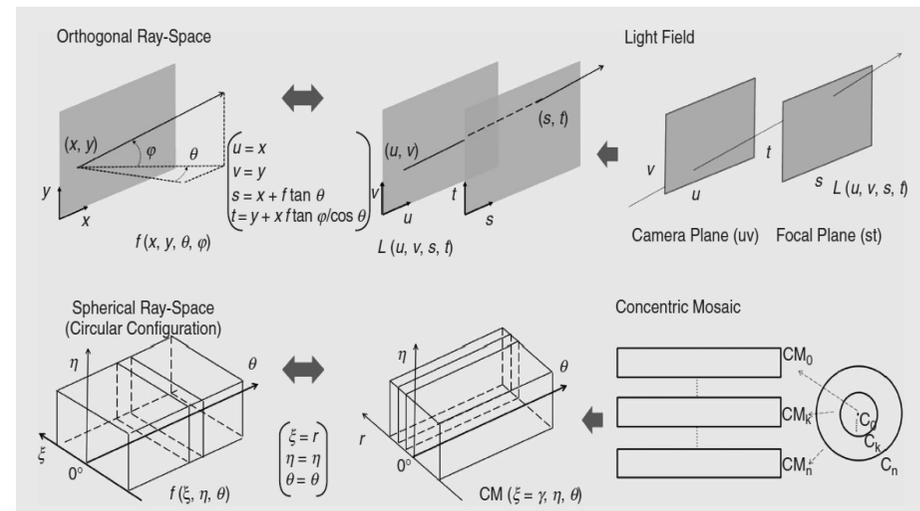


図6 光線空間と Light Field, Concentric Mosaic との関係⁴⁾

一方の平面を無限遠に持っていった場合に直交座標光線空間に一致するとみなすことができる。Light FieldはCGにおけるレンダリングを目的としており、2枚の平面を観察者視点に応じて適切に設定したり、複数の平面の組を使うことで広い視域に対応することができる。一方の直交座標光線空間はレンティキュラーやパララックスバリアなどの3次元画像ディスプレイに表示する共通データフォーマットとして導出された経緯があるため、基準面をそのままディスプレイ面とみなすことができ、3次元ディスプレイの表現方法と整合性がよい。同様の手法としてLumigraph¹⁶⁾、また光線を定義する基準面を物体の大まかな形状と一致させたSurface Light Field¹⁷⁾がある。また、1999年にShumらによってConcentric MosaiCs¹⁸⁾が提案されている。こちらは極座標光線空間と等価である。図6にこれらの関係を示す。

3. 光線空間法の応用

3.1 自由視点映像生成

前章で述べたように、ピンホールカメラでシーンを撮影した際に得られる画像は4次元光線空間の2次元部分空間で与えられる。これを利用して、多視点画像の取得から自由視点画像の生成に至る入出力関係を考えてみる。

まず多視点映像とは、時々刻々変化する光線空間 $f(x, y, \theta, \phi)$ を多数の視点位置において観測したものである。縦方向の視差 ϕ を無視し、 y 方向に積層した3次元光線空間 $f(x, y, u)$ について考える。図7(左)に示すように実空間中に並んだ多視点カメラで3次元シーンを撮影したとする。カメラで撮影した画像は (x, u) 平面上における直線が y 方向に積み重なったものであり、光線空間 $f(x, y, u)$ を $x-u$ 平面に垂直な平面で切った断面に相当するものとなる。図7(左)において、カメラ1, ..., 5で撮影した画像は、光線空間の断面1, ..., 5に相

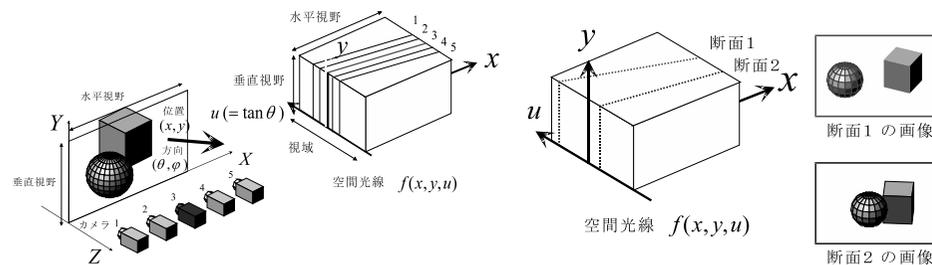


図7 光線空間の取得(左)と自由視点映像の生成(右)

当する。このように、多視点画像の撮影とは、光線空間の断面を観測・記録していることに他ならない。一方、このようにして撮影した画像群をもとに新しい視点の画像を生成するためには、図7(右)に示すようにその視点に対応する新しい断面で光線空間を切り出せばよい。すなわち、多視点画像を入力として自由視点画像を生成するためには、光線空間データの対応する2次元断面を多数観測して光線空間を構成し、新しい視点位置に対応する2次元断面を切り出すことにより実現される。一方、極座標光線空間の場合には、sinカーブの断面を観測・記録していることに相当し、新たな視点の画像生成は別のsinカーブ断面を切り出すことにより行われる。これらの関係を図8に示す。

3.2 可変焦点画像の生成

光線空間の別の応用として、可変焦点画像の生成が挙げられる¹³⁾。図9を参照しながらその原理を述べる。いま、点 $P(X, Z)$ に視点があるものとし、そこに有限の大きさの開口(aperture)を持った薄肉凸レンズを仮定する。ピンホールカメラモデルでは、ピンホール $P(X, Z)$ を通る光線群のみに着目するため、画像面上で得られる画素の輝度値は光線 a の輝度値であって、それは光線空間の対応する a 点の値を取り出せば良いのであった。一

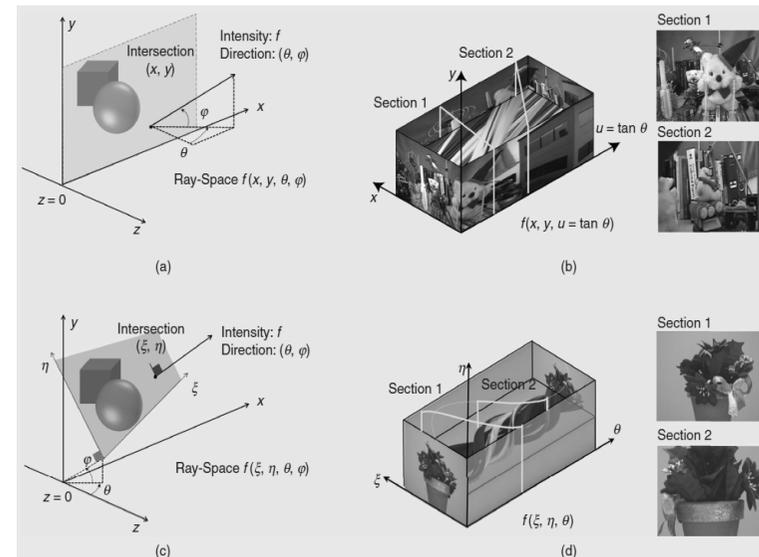


図8 光線空間と自由視点映像生成の例(上:直交座標, 下:極座標)⁴⁾

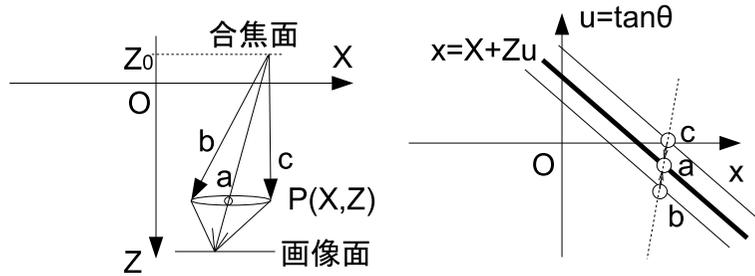


図 9 可変焦点画像の生成

方、図のようにレンズの開口がある場合には、光線 b , c の間のある光線の輝度値が足し合わされて観測されると考える。すなわち、開口の大きさに応じて光線空間内の直線 $X = x + Zu$ の上下に帯ができ、その帯内のデータを足し合わせて観測することになる。足し合わせる方向は合焦面の奥行きによって定められ、合焦面が $Z = Z_0$ の場合に、光線空間内での傾きは $-1/Z_0$ で与えられる。合焦面が物体表面にある場合には、元の光線空間の直線構造と同じ方向に足し合わせることとなり、合焦面が物体表面からはずれた場合には、さまざまな傾きの直線を横断して足し合わせることになる。このようにして、物体が合焦面上からはずれる度合いに応じて「ボケ」として観測されるようになる。

これを応用すれば、撮影時に何らかの方法で 4 次元光線空間を取得しておいて、撮影後に自由に焦点を変化させることができるようになる。文献¹⁹⁾では、後述の IP 方式と同様なマイクロレンズ光学系を用いて 4 次元光線空間を 2 次元画像に変換して取得し、後段の処理で焦点を変化させる機能を実現している。

4. 3 次元映像の取得・表示

3D 映画の成功に牽引された近年の 3D プームに乗って、各社から 3D 対応のテレビや携帯電話、ゲーム機などが発売されている。一方、業務用のステレオカメラやステレオ画像が取得できる携帯電話なども利用可能になりつつある。さらに、視点数も 2 つだけでなく多視点のディスプレイも開発が進み、多視点カメラも出てきている。本章ではこれらを光線空間の入力・出力装置という観点から捉え直し、研究の動向、今後の方向性について論じる。

4.1 取得システム

4.1.1 多視点カメラによる取得

動的シーンの多視点映像を取得するシステムとしては、ビデオカメラを並べる方式が一般的である。カメラ台数が少ない場合には IEEE1394 カメラ等を用いることで、比較的容易に多視点カメラシステムを構築することができる。数十眼以上の大規模システムになると、各カメラの同期を取ることで、および膨大なデータ量となる映像信号を効率良く扱う手法が鍵となる。代表的なシステムとしては、米国スタンフォード大のシステム²⁰⁾ や米国カーネギーメロン大 (CMU) のカメラアレイ²¹⁾ がある。また、名古屋大学では、100 台のカメラアレイを独自に開発し²²⁾、撮影した高解像度映像を MPEG の多視点映像符号化実験に提供している。図 10 に名古屋大学が開発した 100 眼システムの外観を示す。

多視点カメラ方式は、動的シーンの光線取得システムとしては有望であるものの、多数の異なったカメラで撮影した場合、取得した多視点映像間には幾何学的な位置ずれや色や明るさの違いが生じ、これが後の処理・表示において問題となる。従って、取得後に前処理として幾何学的位置合わせや色合わせを行う必要がある。

幾何学的位置合わせ法としては、3次元世界座標が既知の物体をシーン内に置き、その 3次元世界座標と各カメラへの投影点である画像座標との多数の組を元に、投影誤差が最小となるようにカメラパラメータを定めて位置合わせを行う手法が一般的である。これに対して、多視点映像内で対応点が描く軌跡に対して制約を与え、ホモグラフィ (Homography) を用いてカメラの向きや内部パラメータを合わせる手法も提案されている²³⁾。

多視点映像間の色の統一性 (Color Consistency) の確保については、色空間におけるカ

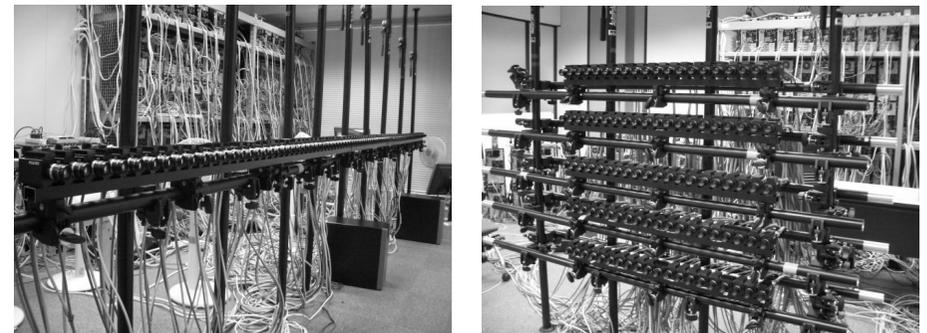


図 10 多視点映像取得システム (名古屋大 100 眼システム²²⁾)

ラーマッチングの手法が適用できる。文献²⁴⁾では、RGBチャンネルごとに独立に輝度の変換関数を想定し、色合わせの基準となる点と対象とする点の色とが同じになるように変換関数の形を求めていく手法をとっている。

このように多視点カメラ方式は動的な光線空間データを取得するために有望な手法ではあるが、カメラ間の位置合わせや色の統一性の確保の問題を解決する必要がある。また光線空間取得の観点からは、多視点カメラで得られる映像データは視点間隔が非常に疎（スペース）であり、光線空間の補間処理が必須となる。これについては今後の課題の項で述べる。

4.1.2 光線空間直接取得法

多視点カメラ方式の発展形として、光線空間データを直接取得するシステムがある。光線空間直接取得法は、大きく二つの手法に分けられる。一つは、光線空間データをレンズアレイ等によって空間分割して2次元データに展開し2次元カメラで撮影する手法であり、例としてNHKが開発したGRINレンズを用いたインテグラルフォトグラフィ（Integral Photography; IP）の直接撮像・表示装置²⁵⁾が挙げられる。もう一つの手法は、高速な撮像素子と時分割を用いて光線空間を取得する手法であり、走査光学系と高速度カメラを用いた手法が例として挙げられる²⁶⁾。近年、このように光学系+撮像素子+信号処理を組み合わせたものが盛んに研究されつつあり、Computational Camera, Computational Photographyという分野へと発展をみせている。これらは光線の取得という観点からは非常に興味深いアプローチであり、今後さらなる研究の発展を期待したい。

4.2 3次元映像表示

3次元映像表示を論ずる際、どのような仕組みで3次元表示を可能にしているか、という点に興味が集まりやすい。例えば、メガネを使う方式では、光の波長（赤青メガネ；アナグ

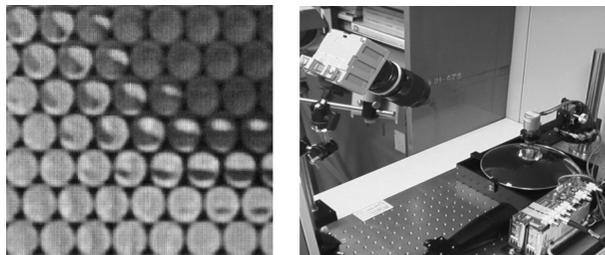


図 11 左：空間分割による光線取得（Integral Photography; IP, NHK 技研提供）、右：時間分割による光線取得（走査光学系と高速度カメラ）

リフ）や偏光方向（直線・円偏光）による左右像の分離や、時分割（アクティブ・シャッター方式）による左右像の分離などである。またメガネなしの方式では、レンチキュラーレンズを使うものやパララックス・バリア方式などが挙げられる。一方で、これらを光線空間の表示装置と捉えるならば、視点数がいくつか、すなわち光線空間の2次元断面が何枚必要か、という点のみが問題である。自由視点テレビ（FTV）³⁾では、単一視点の映像をユーザが切り替えながら見るという機能を提供しているが、この場合に切り出す断面数は1である。メガネ方式では、視点数は2であり、観察者の位置によらず切り出す断面は固定である。メガネ方式でもアイトラッキング等を行って視点追従の機能を持たせるならば、インタラクティブに切り出す2つの断面を変える必要がある。多視点ディスプレイになると、切り出す断面も視点数に応じて増やす必要がある。いずれにしても、平面3次元ディスプレイであれば直交座標光線空間のデータが利用可能であればよく、視点数に応じて光線空間の視点(u)方向の「解像度」が異なるだけという見方ができる。

また近年、新しいタイプの3次元ディスプレイが話題となっている。ソニーの「360°立体ディスプレイ」や、NICT けいはんな研究所ユニバーサルメディア研究センターの「fVisiOn」（エフ・ビジョン）など、全周から観察可能な立体ディスプレイがその一つである（図12）。これらは、極座標光線空間を表示できるデバイスと考えることができ、今後は物体を取り囲むような全周囲の多視点画像を対象とした研究も盛んになってくるであろう。

5. 課題と今後の展望

5.1 光線空間の補間

光線空間データが得られれば、3次元映像・自由視点映像生成は視点位置に対応する断面



図 12 左図：「360°立体ディスプレイ」（ソニー）、右図：「fVisiOn」（NICT）

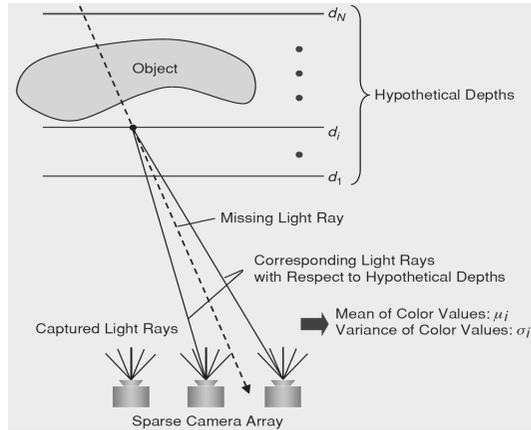


図 13 光線の補間¹⁾

を切り出すことにより実現される。しかしながら、通常が多視点カメラから得られる光線のサンプルは非常に疎な間隔で観測されたものである。どれだけの数のサンプルがあれば、原光線空間を再構成できるかという問題は、Plenoptic Sampling²⁷⁾の理論として知られており、光線空間の周波数的解析によって Lambert 表面仮定上で隣接視差画像間での視差が 1[pixel] 以内であることを要請するとしている。この値を満たすためには、カメラを非常に密に配置する必要があり、カメラの物理的大きさを考慮すると現実的ではない。従って、図 13 に示すように取得されていない光線をいかにして「補間」するか、という技術が鍵となる。この光線を補間する手法としては、撮影対象の幾何モデルを陽に用いるか否かを大まかな基準として、幾何モデルベース法 (Model-based Rendering, または Geometry-based Rendering と呼ばれる) とイメージベース法 (Image-based Rendering) とに大別されて議論されることが多い。しかしながら、幾何モデルを利用する度合いに応じて両者の中間的に位置する手法はさまざまに考えられる。図 14 は横軸方向に幾何モデルを利用する度合いをとり、現在までに提案されているさまざまなレンダリング手法をその軸上に位置づけたものである。詳細は文献²⁸⁾を参照されたい。

光線補間の一例として、文献²⁹⁾で提案した手法を紹介する。いま、 N 点の視点で撮影した多視点画像列 $I_n(i_n, j_n) (n \in \{1, 2, \dots, N\})$ を考える。そして、3次元空間内の点 (X, Y, Z) を通る光線が、多視点画像 I_n 内の画素 (i_n, j_n) で観測されたとする。カメラ n の透視投影

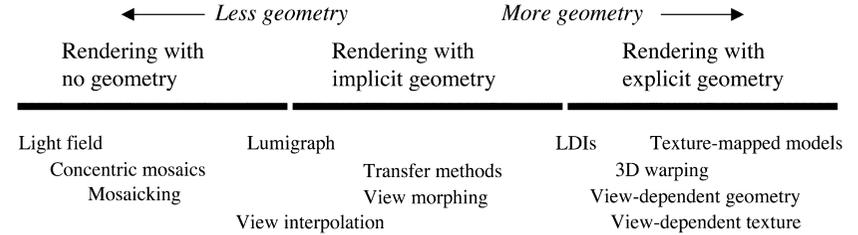


図 14 自由視点映像生成手法の分類²⁸⁾

行列を $P_n (\in R^{3 \times 4})$ としたときに、 (X, Y, Z) と (i_n, j_n) は同次座標を用いて以下のように関係づけられる。

$$\begin{pmatrix} i'_n \\ j'_n \\ s \end{pmatrix} = P_n \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix}, \quad i_n = i'_n/s, j_n = j'_n/s. \quad (6)$$

いま P_n が既知であるとする、 (X, Y, Z) が与えられると (6) 式により (i_n, j_n) が一意に定まる。このときの画像 $I_n(i_n, j_n)$ の画素値を L_n と書くと、 L_n は (X, Y, Z) の関数となる。すなわち、3次元空間中の各点 (X, Y, Z) に対して、この点を通る N 本の光線の輝度値 L_n が観測されたと考えるのである。これらに対し、以下の量を考える。

$$Avg(X, Y, Z) = \frac{1}{N} \sum_{n=1}^N L_n(X, Y, Z), \quad (7)$$

$$Var(X, Y, Z) = \frac{1}{N} \sum_{n=1}^N \{L_n(X, Y, Z) - Avg(X, Y, Z)\}^2. \quad (8)$$

$Var(X, Y, Z)$ は、 (X, Y, Z) を通る N 本の光線の輝度値の分散を表しており、Lambert 反射則に従う物体表面が存在する場合には、 $Var(X, Y, Z)$ の値が小さくなる。 $Avg(X, Y, Z)$ は N 本の光線の輝度値の平均値である。これらをそれぞれ分散空間、平均の色空間と呼ぶ。この 2 つの空間を用いて、ワイヤフレームモデルの表面に沿った $Var(X, Y, Z)$ の総和が最も小さくなるように形状を定め、定まった形状表面に $Avg(X, Y, Z)$ の値を貼っていくという手法に基づいて物体の形状抽出を行った例を図 15 に示す。この例では、平均と分散を用い

たが、視点の近さ等に応じて $L_n(X, Y, Z)$ に重みをつけて総和をとってもよいし、対応点の類似度合いを表す関数として他の関数を用いてもよい。類似の手法として Plane Sweeping 法³⁰⁾があり、これは各多視点画像を 3次元空間中の仮想平面上に射影して類似度を考え、その仮想平面を奥行き方向に走査させて行く手法である。

また近年、コンピュータビジョンの分野でステレオ視問題をエネルギー最小化問題としてとらえ、Graph-Cut などのアルゴリズムによって解く手法が盛んに研究されている。このエネルギー関数は次のような形で与えられる³¹⁾。

$$E(L) = \sum_{p \in P} D_p(L_p) + \sum_{(p,q) \in N} V_{p,q}(L_p, L_q) \quad (9)$$

L は画素に割り当てるラベル（この場合、奥行き値）である。 D_p はデータ項と呼ばれ、対応点が存在する確からしさを表す。 $V_{p,q}$ は平滑化項と呼ばれ、近傍画素間での奥行き値が滑らかに変化するという拘束条件となる。Graph-Cut の詳細は文献³²⁾等を参照されたい。これらは主にステレオ視での研究であるが、多視点画像が利用できる場合には、データ項に多視点画像を利用した対応点存在尤度を入れることができるため、より安定に奥行き推定を行うことができるようになる。

この他に興味深い手法としては、物体の奥行きを一つに定めず、信頼度マップと呼ばれる確率モデルを用いたレンダリング手法³³⁾や、信号処理的な手法によって視点内挿を行う手法³⁴⁾などが提案されている。また最近、超解像を使った自由視点映像の高解像度化の研究が行われている^{35),36)}。超解像技術が一般的になっている現在、自由視点映像の新しいレンダリング手法として今後の発展が期待される。

5.2 光線空間の情報圧縮

光線空間は非常に情報量が大きく、2次元映像の数百～数千倍に及ぶこともある。一方で、そのデータは非常に冗長であり効率のよい情報圧縮手法が検討できる。Light Field

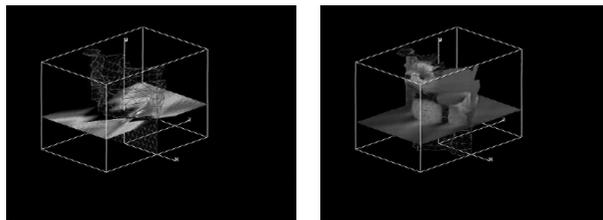


図 15 分散空間、平均の色空間を用いた形状抽出^{2),29)}

Rendering¹⁵⁾では、ベクトル量子化 (VQ) と Lempel-Ziv(LZ) 符号化を用いて圧縮を行っていた。一方で、光線空間はその定義から「微小にずらした位置から撮影した多数の 2次元映像の集まり」と考えることができることから、2次元画像の情報圧縮における長年の研究成果を応用することができる。その代表例が、MPEG-2 や H.264/AVC で使われている予測符号化と変換符号化 (DCT など) を組み合わせた方式である。それをさらに発展させて 3次元映像、さらには光線空間の符号化方式へと応用が可能である。このような考えのもと、自由視点テレビおよび光線空間の符号化方式を国際標準化組織である MPEG に提案し、標準化活動を行った。以下、MPEG における 3次元映像符号化の活動について概観する。

MPEG における 3次元映像符号化は、2001年12月に行われた日本の JEITA からの 3D Video 標準化提案をきっかけとして開始された。その初期の頃からの標準化活動については文献³⁷⁾を参照されたい。その後、光線空間符号化を考える前に、まず多視点映像をそのまま符号化・伝送できる方式を策定しようという機運が高まり、多視点映像符号化 (Multiview Video Coding; MVC) の技術提案募集が 2005年7月に行われた。その後の 2006年1月の会合において各方式を比較検討した結果、H.264/AVC をベースとして予測を視差軸方向に拡張した方式が採用された。基本的な圧縮技術としては、動画像符号化で用いられている時間方向の予測 (動き補償) を多視点画像の視差画像 (view) 間の予測として用いるものであり、視差補償予測符号化方式と呼ばれるものの一形態と位置づけることができる³⁸⁾。図 16 に MVC における視差補償予測方式の概念図を示す。2009年には H.264/AVC Annex H: Multiview Video Coding (MVC) として規定された。2009年12月に Blu-Ray Disc

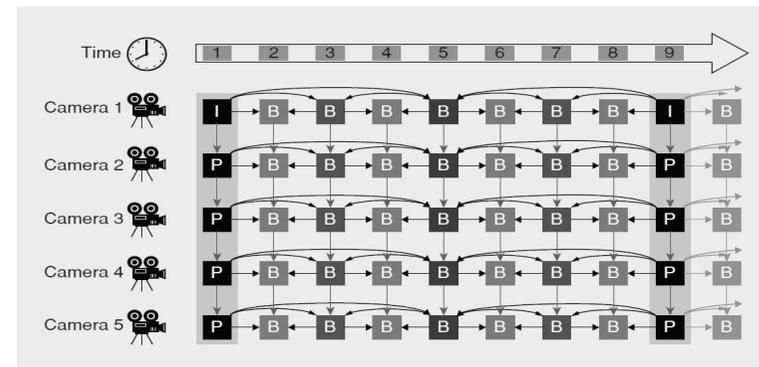


図 16 多視点映像符号化 (MVC) における予測方式¹⁾

Association(BDA)「Blu-ray 3D」の仕様として MVC を採用するなど、普及が進みつつある。

視差画像 (View) に加えて奥行きマップ (Depth) を補助情報として加えることにより、視域が狭い範囲であれば CG 的なレンダリング手法を用いて多視点映像を生成できる。この考え方に基づいた方式が 2005 年 10 月にフィリップスから提案され、その後の標準化作業を経て 2007 年に ISO/IEC 23002-3 (MPEG-C Part 3) として規定された。図 17 に View+Depth フォーマットの一例を示す。また、レイヤー (層) ごとに情報を持っているデータ形式からレンダリングする手法として Layered Depth Image³⁹⁾ が提案され、これを動画像に拡張した Layered Depth Video(LDV) が検討されている。また、前述の View+Depth, MVD の形式では、視差画像と depth 情報を用いて自由視点の映像を生成する方式が検討されており、Depth-Image-Based-Rendering と呼ばれている⁴⁰⁾。各視点からの depth 情報しか与えられていないため、ワーピングにより視点変換画像を合成した場合には depth が急激に変わるエッジ付近等で生成画像に欠落が生じる。この品質劣化を軽減するための研究が行われている^{41),42)}。

現在、表示側においてディスプレイの仕様に合わせてベースライン長を変更したり、視差画像を新たに生成して眼数の異なる多視点ディスプレイへの表示も可能な方式として、3D 映像符号化 (3D Video Coding: 3DVC) の検討が行われている⁴³⁾ (図 18)。ここでは、視差画像 (view) の情報に加えて奥行き (depth) 情報を付加したものが提案されており、View+Depth, Multi-View+Depth(MVD) といった方式が検討されている。現在、奥行き推定 (Depth Estimation)、視差画像生成 (View Synthesis) の参照ソフトが規定され、符号化実験が続けられているところであり、2011 年初頭に Call for Proposals が発行されるスケジュールで検討が進められている。生成画像の品質の向上という観点から視差画像

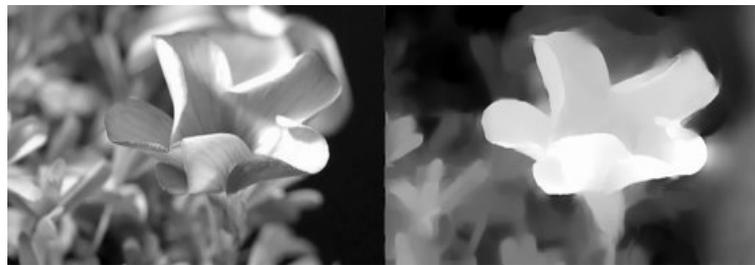


図 17 View + Depth フォーマット (Philips の Web site より)

と depth 情報をどのように圧縮・符号化していくかという研究が今後盛んになっていくことが予想される。

6. ま と め

本稿では、3次元映像処理の基礎となる概念である光線空間の概要を紹介し、自由視点映像の生成、可変焦点画像の生成への応用、さらに3次元映像の取得・表示手法との関連について述べた。

近年の3次元映像に関する関心の高まりを受け、さまざまな機関で興味深い研究が行われている。情報通信研究機構 (NICT) では、平成 21 年度から「革新的な三次元映像技術による超臨場感コミュニケーション技術の研究開発」の委託研究を開始しており、各機関で三次元映像通信の実現に向けた要素技術の研究がなされている⁴⁴⁾。その他、NHK 放送技術研究所では伝統芸能の3次元映像アーカイブ化技術を開発しており、KDDI 研究所では多視点映像を元に、シリンダー空間分割法により被写体の間に入り込んだ視点からの映像 (ウォークスルー映像) の合成を行っている。これらは視体積交差法やマッチング処理、5章で紹介した Graph-Cut などを組み合わせて自由視点映像を生成しており、今後とも同様のシステムの開発が次々と進むであろう。

これらの大規模なプロジェクトの根底には、光線空間の基本的な取得、サンプリング・補間問題、情報圧縮の問題が横たわっており、これらをじっくりと研究していくことが求められる。さらに、多視点画像の幾何学的キャリブレーション、奥行き・形状推定、3次元映像

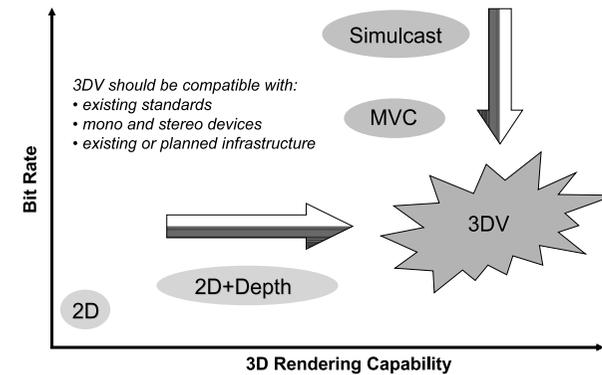


図 18 Vision on 3D Video⁴³⁾

符号化, 3次元映像・自由視点映像生成等の研究においても, キャリブレーションの簡易化, 奥行き推定の精度向上, depth情報の符号化, 高品質な自由視点映像生成などのさまざまな研究課題があり, 今後ますますの研究の発展が期待される。

参 考 文 献

- 1) IEEE Signal Processing Magazine, Special Section – Multiview Imaging and 3DTV – vol. 24, no. 6 (2007.11).
- 2) IEEE Trans. CSVT – Special Issue on Multiview Video Coding and 3DTV –, vol. 17, no. 11 (2007.11).
- 3) 谷本, “自由視点テレビFTV,” 信学論 (A), vol. J89-A, no. 11, pp. 866–872, (2006.11).
- 4) M.Tanimoto, M.P.Tehrani, T.Fujii, and T.Yendo, “Free-Viewpoint TV,” IEEE Signal Processing Magazine, pp. 67–76, January (2011).
- 5) “Vision on 3D Video,” ISO/IEC JTC1/SC29/ WG11 N10357 (2009.2).
- 6) 原島, 藤井, 苗村, “3次元統合画像通信へむけて,” PCSJ '92 (1992.10).
- 7) E.H. Adelson and J.R. Bergen, “The Plenoptic Function and the Elements of Early Vision,” chapter 1 in Computational Models of Visual Processing, Cambridge, MA: MIT Press, pp. 3–20 (1991).
- 8) T. Fujii H. Harashima, “Coding of an Autostereoscopic 3-D Image Sequence,” SPIE VCIP, vol. 2308, pp. 930–941 (1994.9).
- 9) 藤井, 3次元統合画像符号化の基礎検討, 東京大学工学系研究科, 博士論文, 1995. (<http://hdl.handle.net/2261/1866>)
- 10) 藤井, 原島, “構造抽出に基づく三次元画像符号化,” 画像符号化シンポジウム (PCSJ94), pp.23–24 (1994.10).
- 11) T. Fujii, T. Kimoto, M. Tanimoto, “Ray Space Coding for 3D Visual Communication,” Picture Coding Symposium '96, pp. 447–451 (1996.3).
- 12) T. Fujii and M. Tanimoto, “Freeviewpoint TV system based on Ray-Space Representation,” SPIE ITCOM 2002, vol. 4864-22, pp. 175–189 (2002.8).
- 13) 藤井, 金子, 原島: “光線群による3次元空間情報の表現とその応用,” テレビジョン学会誌, vol.50, no.9, pp. 1312–1318 (1996.9).
- 14) Robert C. Bolles, H. Harlyn Baker, and David H. Marimon, “Epipolar-Plane Image Analysis: An Approach to Determining Structure from Motion,” International Journal of Computer Vision, 1, pp. 7–55 (1987).
- 15) M. Levoy and P. Hanrahan, “Light Field Rendering,” Proc. ACM SIGGRAPH '96, pp. 31–42 (1996.8).
- 16) S. Gortler, R. Grzeszczuk, R. Szeliski, and M. Cohen, “The Lumigraph,” Proc. ACM SIGGRAPH '96, pp. 43–54 (1996.8).
- 17) Daniel N. Wood, Daniel I. Azuma, Ken Aldinger, Brian Curless, Tom Duchamp, David H. Salesin and Werner Stuetzle, “Surface Light Fields for 3D Photography,” in Proc. ACM SIGGRAPH 2000, pp. 287–296 (2000.8).
- 18) H. Y. Shum and L. W. He, “Rendering with concentric mosaics,” in Proc. ACM SIGGRAPH '99, pp. 299–306 (1999.8).
- 19) Ren Ng et al., “Light Field Photography with a Hand-held Plenoptic Camera,” Stanford Tech Report CTSR 2005-02 (2005).
- 20) B.Wilburn, et al., “High Performance Imaging Using Large Camera Arrays,” Proc. ACM SIGGRAPH 2005 (2005.8).
- 21) C.Zhang and T. Chen, “A Self-Reconfigurable Camera Array,” Eurographics Symposium on Rendering 2004 (2004.6).
- 22) T. Fujii, K. Mori, K. Takeda, K. Mase, M. Tanimoto, Y. Suenaga, “Multipoint Measuring System for Video and Sound: 100-camera and microphone system,” IEEE ICME 2006, pp. 437–440 (2006.7).
- 23) 福嶋, 松本, 圓道, 藤井, 谷本, “特徴点軌跡並行化による2次元カメラアレイレクティファイケーション,” 映像情報メディア学会誌, vol. 62, no. 4, pp. 564–571 (2008.4).
- 24) 山本, 圓道, 藤井, 谷本, “Colour Correction for Multiple-camera System by using Correspondences,” 映像学誌, vol. 61, no. 2, pp. 213–222 (2007.2).
- 25) 洗井, 岡野, 星野, 湯山, “屈折率分布レンズを用いたインテグラルフォトグラフィの撮像方式の基礎検討,” 映像学誌, vol. 51, no. 11, pp. 1927–1934 (1997.11).
- 26) 藤井, 谷本, “走査光学系と高速度カメラを用いた光線空間のリアルタイム取得装置,” 3次元画像コンファレンス 2004, pp. 197–200 (2004.6).
- 27) Jin-Xiang Chai et al., “Plenoptic Sampling,” Proc. ACM SIGGRAPH 2000, pp. 307–318, (2000.8).
- 28) H.-Y. Shum, S. Kang, S.-C Chan, “Survey of Image-Based Representations and Compression Techniques,” IEEE Trans. CSVT, vol. 13, no. 11, pp. 1020–1037 (2003.11).
- 29) T. Fujii and H. Harashima, “Data Compression and Interpolation of Multi-View Image Set,” IEICE Trans. on Inf. & Sys., vol. E77-D, no.9, pp. 987–995 (1994.9).
- 30) R.T. Collins, “A Space-Sweep Approach to True Multi-Image Matching,” Proc. IEEE CVPR'96, pp. 358–363, 1996.
- 31) Y. Boykov, V. Kolmogorov, “An Experimental Comparison of Min-Cut/Max-Flow Algorithms for Energy Minimization in Vision,” IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI), Vol. 26, No. 9, pp.1124–1137 (2004.9).
- 32) 石川: “グラフカット,” 情報処理学会 CVIM 研究会チュートリアル IPSJ SIG Notes. CVIM vol.2007, no.31, pp. 193–204 (2007.3).
- 33) 國田, 上野, 田中, “多層信頼度マップを用いた3次元映像の実時間生成システム,” 映像学誌, vol. 60, no. 7, pp. 1102–1110 (2006.7).
- 34) 久保田, 児玉, 羽鳥, “奥行き推定を必要としない視点内挿手法とその安定性解析 –円

- 周配置の多視点画像を用いた中心視点画像の生成-,” 信学論 (D) , vol. J90-D, no. 4, pp. 1063–1072 (2007.4).
- 35) 福嶋, 石橋, “超解像処理による自由視点画像の画質改善,” 信学論 (D), vol. J93-D, no.9, pp.1700–1703 (2010.9).
- 36) 石井, 高橋, 苗村:“超解像的アプローチによる視点補間の基礎検討,” 信学論 (D), vol. J93-D, no.9, pp.1682–1684 (2010.9).
- 37) 木全, “MPEG 3DAV 国際標準化の動向,” 映情学誌, vol. 60, no. 2, pp. 143–149 (2006.2).
- 38) K. Muller, P. Merkle, T. Wiegand, “Compressing Time-Varying Visual Content,” IEEE Signal Processing Magazine, vol. 24, no. 6, pp. 58–65 (2007.11).
- 39) J. Shade, S. Gortler, L He, R. Szeliski, “Layered Depth Images,” ACM SIGGRAPH ’98 (1998.7).
- 40) C. Fehn, “Depth-Image-Based Rendering (DIBR), Compression and Transmission for a New Approach on 3D-TV,” Proceedings of SPIE Stereoscopic Displays and Virtual Reality Systems XI, pp. 93–104 (2004.1).
- 41) K. Muller, A. Smolic, K. Dix, P. Merkle, P. Kauff, and T. Wiegand, “View Synthesis for Advanced 3D Video Systems,” EURASIP Journal on Image and Video Processing, Special Issue on 3D Image and Video Processing, vol. 2008, Article ID 438148 (2008).
- 42) L. Yang, T. Yendo, M. P. Tehrani, T. Fujii, M. Tanimoto, “Probabilistic Reliability Based View Synthesis for FTV,” IEEE International Conference on Image Processing 2010, pp. 1785–1788 (2010.9).
- 43) ISO/IEC JTC1/SC29/WG11 N10357, “Vision on 3D Video” (2009.2).
- 44) <http://www.nict.go.jp/>