

# 平面群構造を利用した車載カメラ映像からの オンライン前景障害物除去システム

栗林宏輔 \*<sup>1</sup> 子安大士 \*<sup>1</sup> 前川仁 \*<sup>1</sup> 小野晋太郎 \*<sup>3</sup> 川崎洋 \*<sup>2</sup> 池内克史 \*<sup>3</sup>  
埼玉大学 〒 338-8570 埼玉県さいたま市桜区下大久保 255\*<sup>1</sup>  
鹿児島大学 〒 890-0065 鹿児島市郡元 1-21-40\*<sup>2</sup>  
東京大学 〒 153-8505 東京都目黒区駒場 4-6-1\*<sup>3</sup>

都市モデリングをする際の車載カメラで撮影した映像には得たい都市の建造物や道路などの背景以外にも歩行者や街路樹などの前景物が映り込むことがよくある。これらはモデリングの際に障害になるなど、画像上から除去する必要がある。そこで本論文では、背景に遮蔽が起こっている車載カメラ映像から自動的に背景のみの映像を作成する手法を提案する。具体的には、建造物などの平面性を利用して背景をトラッキングする。次に、時空間画像解析によって背景色を推定し、遮蔽が起こっている領域にその色を上書きすることで前景物を画像上から除去する。複数の背景平面に対応するオフラインシステムと単一平面に対応するオンラインシステムの両方を実装し、それぞれの実験結果を示す。

## Online System of Foreground Obstacle Removal from On-Vehicle Video Using Plane Structure Constraints

Kousuke KURIBAYASHI\*<sup>1</sup> Hiroshi KOYASU\*<sup>1</sup> Hitoshi MAEKAWA\*<sup>1</sup> Shintaro ONO\*<sup>3</sup> Hiroshi  
KAWASAKI\*<sup>2</sup> Katsushi IKEUCHI\*<sup>3</sup>  
Saitama University 255 Shimo-Okubo, Sakura-ku, Saitama City, Saitama 338-8570, JAPAN\*<sup>1</sup>  
Kagoshima University 1-21-40 Kagoshima Kourimoto, JAPAN\*<sup>2</sup>  
University of Tokyo 4-6-1 Komaba Meguro-Ku, Tokyo 153-8505, JAPAN\*<sup>3</sup>

**Abstract** In the urban scene modeling, removing foreground obstacles (e.g. pedestrians, telegraph poles, roadside trees, etc.) is important problem for 3D reconstruction, privacy protection, etc. We propose the method which can automatically remove foreground objects for such problem from the on-vehicle video data. Since an input data is a video stream and an urban scene is mainly composed of planar surfaces the method can effectively separate the objects by using the Spatio-temporal image analysis. To show the effectiveness of the method, several experiments using real data are conducted, which resulted in a successful separation of complicated objects.

**Keyword:** *Spatio-temporal analysis, structure from motion, obstacle removal*

### 1. はじめに

都市空間をカメラや各種センサによって計測し、モデリングしようとする試みが近年非常に盛んである。特に車にカメラを載せ、街の中を走りながら撮影する手法は数多く提案されている。しかし、そうした車載カメラによって撮影した映像中には得たい建造物や道路などの街の景観だけでなく、歩行者や街路樹などの恒常

的に存在しない物体が映り込み、モデリングの際の障害となることがよくある。このように、撮影した画像上に撮影対象物とそれを遮蔽する物体が同時に映りこむ状況は日常的に見られる現象である。この現象は撮影者の意図通りの画像取得を妨げるだけでなく、Google Earth や Virtual Earth などでは誤ったテクスチャマッピングの原因となり、Intelligent Transport Systems に

おける標識認識や歩行者、駐車車両検出の際にも大きな問題 となっている。また、最近では Google Street View などのシステムにおいて、映り込んでしまった人や自動車などによるプライバシー侵害の問題なども引き起こしている。

そこで、この問題の解決策として画像上から撮影対象物であるビルや道路などの常に存在する物体を背景、背景より手前に存在する物体を前景障害物（前景物）と定義し、前景物により遮蔽が起こったシーンから前景物領域の奥に存在する背景の色情報を求め、その領域に上書きすることで前景物を除去する手法を提案する。

主な手段として時空間画像解析を用いる。時空間画像解析では例えばカメラが等速直線運動をしている場合の EPI<sup>1</sup> 上では現実世界のシーンにおける 3 次元点が直線となって現れ、直線上での補間処理によって、複雑な形状をした前景物や移動物体や複数個の集合であっても、容易に除去することができる。

本論文では、背景が複数の平面で構成されたシーンに対応するオフライン処理と背景が一枚平面で構成されたシーンに対応するオンライン処理の両方を提案する。オフライン処理による実験では、実際にカメラで市街のデータを撮影し、建物や街路樹を撮影するにあたり前景物と考えられる対象を自動で除去することが出来た。オンライン処理による実験では、USB カメラで撮影し、約 0.6fps の処理速度によって前景物を除去することが出来た。そこで、本論文では、まず最初に関連研究について述べ、次にオフライン処理による提案手法全体についてとオンライン処理可能な実装についての詳細を説明し、続いて実験およびその結果を紹介し、最後に結論と考察を述べる。

## 2. 関連研究

固定カメラによる背景作成の研究などは、時間的な補間による前景物除去と考えることが出来る。このような研究はこれまで非常に多く行われている。例えば、Korah らはモーションキューを使用し周りから補間する方法を提案した<sup>2)</sup>。また、Matsushita らは、屋外に固定カメラで撮影した交差点映像から光源（太陽）の動きを考慮することで、車両などの前景物を効率よく除去する手法を提案している<sup>3)</sup>。これらの手法では、カメラが固定され静止しているため、動いている物体を前景物として自動で認識し除去することが可能であり、歩行者などの除去には適しているが、フェンスや街路樹などの除去は難しい。さらにカメラが動く場合に適用することは容易ではない。

前景物に限らず、空間的な画像や映像上の特定の領域を違和感無く穴埋めすることは、需要も大きく長く研究されてきた。最近ではグラフィクス分野において、Li らが lazy snapping による簡易なインタラクティブ操作により、グラフィカットにより高精度に推定し、周辺的数据を用いて滑らかに補間する手法（inpainting）を提案している<sup>1)</sup>。Shiratori らは、ビデオ映像におい

<sup>1</sup>撮影データの時空間ボリュームを各画像のエピポーラ線で切断した面, Epipolar Plane Image

てグラフィカットを用いて、特定の対象物体を除去し時空間内で類似のセグメントを探索することで、滑らかにビデオ映像を補間する手法を提案した<sup>4)</sup>。これらの手法は、前景物が 1 つの大きな塊である場合には極めて有効と考えられるものの、提案手法で想定するような街路樹や通行人など、前景物が複数個や不特定多数個の場合には一つ一つの領域を特定するのは難しい。

時空間的な除去として、羽下らは、時空間画像解析を行い、前景物領域をエッジなどから抽出し、背景の傾きを分散で求め、その傾きで前景物領域の端から端へ滑らかに補間することで前景物除去した<sup>5)</sup>。この手法では前景物の判定をエッジ検出により行っているが、前景物のエッジが必ずしも検出できるとは限らず、また、等速直線運動を仮定した手法のため、本論文の目的としている屋外環境での等速直線運動でない場合や前景物の形状が複雑な場合に利用することは難しい。福地らは、固定首振りカメラの画像を時空間解析し、レンズに付いた水滴を除去した<sup>6)</sup>。しかし、本論文の目的としている都市などの広域空間に対しては、このシステムでは撮影できる範囲が限られているため適用困難である。Garg らは、対象物体を様々な角度や照明条件から撮影されたオクルージョンのない画像を選択し、対象物体の形状や反射特性解析を行ってモデルを作成し、オクルージョンのある画像にそのモデルを当てはめ、オクルージョンを除去する手法を提案した<sup>7)</sup>。しかし、オクルージョンの有無を自動で判別することは難しく、本論文の目的とする自動除去は不可能である。

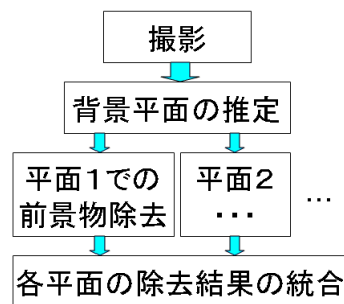


Fig.1 オフライン処理による提案手法全体の処理の流れ

## 3. オフライン処理による提案手法

### 3.1 手法の概要と前提条件

提案手法においては、都市などの広域空間において、建物とカメラの間にある物体を前景物と定義し、これを画像上から除去し、背景の色で補間する。具体的には、前景物とは歩行者や街路樹、電線・電柱などの動物体や街の景観に必要なものではないものである。

また、主なターゲットとして市街地を想定しているため、シーン中で背景として残したい対象としては、建物や道路のように平面で構成される物体か、あるいはカメラの移動距離（ベースライン）に対して十分に遠くに存在する物体とする。

映像はある程度撮影対象物を中心に撮影し、時空間



Fig.2 USB カメラによるオンライン実験結果

画像解析を行うために必要とされるフレームレート以上で撮影されたものとする。また、カメラの運動としては急激な変化は無いこととした。

まず、背景が複数の平面で構成された空間を対象としたオフライン処理について述べる。図 1 に処理の流れを示す。シーンをカメラで撮影し画像列を取得する。各画像から特徴点を抽出し対応点探索を行い、各点の軌跡を求める。軌跡を解析し平面を求め、各平面ごとに前景物除去処理をする。最後にそれぞれの平面の結果を統合し、シーン全体の除去結果を得る。

### 3.2 背景平面の推定

3.3 節の時空間画像フィルタは平面毎にかけるため、背景平面を推定しておく必要がある。

まず、撮影した各入力画像に AffineSIFT<sup>9)</sup> によって特徴点抽出と対応点探索を全画像間で行い、各特徴点の軌跡を求める。AffineSIFT とは Morel らによって提案された SIFT<sup>8)</sup> を元にしたアフィン変化にも頑健な特徴量である。車載カメラによる撮影では撮影対象物を同じ角度から撮影し続けることは難しいが、AffineSIFT を用いることによって撮影位置の移動や撮影角度が変化した場合でもロバストに対応点探索を行える。

求めた特徴点群の軌跡から、RANSAC ベースの射影変換の繰り返しによる平面の検出<sup>11)</sup> を行い、点群を平面ごとにクラスタリングする。平面にクラスタリングされた点群を基に平面の位置合わせを行い、入力画像から射影変換した画像列を生成し、時空間画像フィルタの入力とする。

### 3.3 時空間画像フィルタによる前景物除去

3.2 節で射影変換した画像列を重ね合わせ、時空間ボリュームを構築する。時空間ボリュームの空間的な全ての座標に対して、時間方向の全ての画像を使用し、ベクトルメディアンフィルタ<sup>12)</sup> (VMF) によって背景色を決定する。

時空間ボリュームにおいて前景物がない場合には、ある座標のピクセル値 (RGB の組み合わせ) は背景のある一点が映り続けていることによって全ての画像で同

じとなる。前景物がある場合には、異なったピクセル値が現れるが、背景色が一定のピクセル値で現れ続けるのに対して、異なる撮影位置からの画像を射影変換しているために前景物は背景の前をスライドするように現れ、前景物の色でも同じピクセル値が現れることは稀である。そのため、検出されるピクセル値は背景のものが支配的となり、VMF によって高い確率で背景のピクセル値を取得できる。但し、前景物が一樣な色によって構成されている場合には除去に失敗することもある。VMF の出力  $m$  は次式で表される。 $v$  はピクセル値をベクトル化したものである。

$$m = \underset{v \in \{v_1, \dots, v_N\}}{\operatorname{argmin}} \sum_{i=1}^N |v - v_i| \quad (1)$$

各ベクトルへの距離の和が最も小さいベクトルを選ぶことを意味している。

### 3.4 各平面の除去結果の統合

各平面における除去結果を統合し、シーン全体の除去結果を作成する。しかし、ここまで得ている平面とは特徴点ベースの射影的關係のみであり、平面領域については把握していない。

そこで、各座標ごとに各平面の除去結果を比較し、選択していく。選択方法は VMF によって得た背景色のピクセル値から半径  $r$  以内 (本論文では  $r=2$  とする) のピクセル値の数をカウントし、最も多い平面の背景色を選択する。

## 4. オンライン処理のための実装

本章では §3. で述べた複数の背景平面に対応したオフライン処理に対して、単一平面に対応したオンライン処理への変更点について述べる。

まず、3.2 節の背景平面の推定に用いる特徴量を AffineSIFT から SURF<sup>10)</sup> にすることで計算時間の大幅な短縮を計る。また、1 つ目の平面を検出すると、新たな平面は探索しないようにした。3.3 節の処理について共通である。ただし、撮影を続けることによって貯



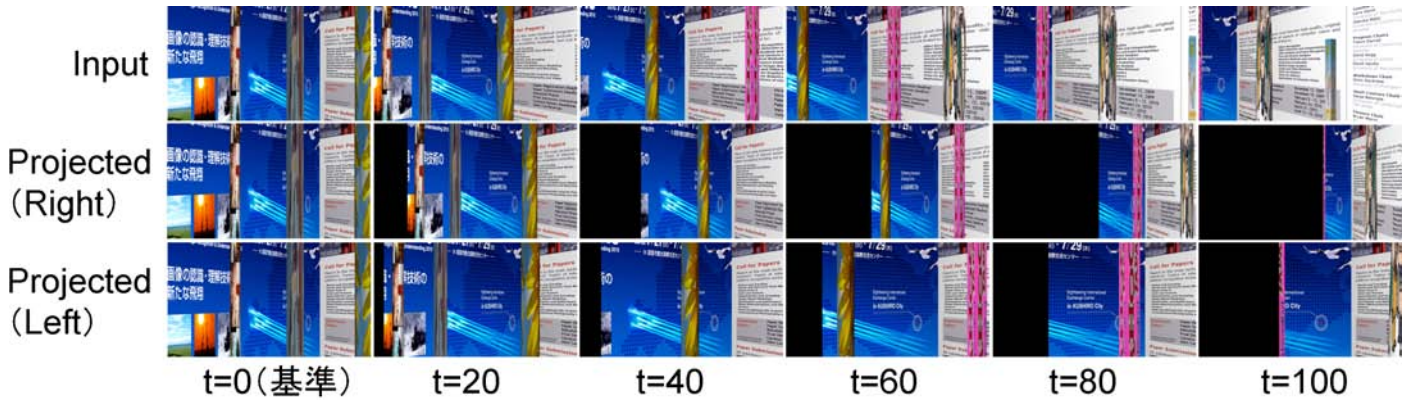


Fig.3 背景が複数平面環境におけるオフライン実験結果

めた画像データを全て時空間画像フィルタの入力とすると処理時間が増していくため、過去  $P$  枚（本論文では  $P = 20$  とする）の画像を使用することとし、一定の fps で処理できるようにした。3.4 節の処理については単一平面のため省略できる。

## 5. 実験

### 5.1 USB カメラによるオンライン実験

USB カメラを手で持ちながら撮影し，§4. で説明したオンライン処理による実験を行った．実験環境は CPU が Intel 製 Core-i7 860 2.80GHz を使用し，画像サイズが  $640 \times 480$  pixel である．実験の結果，処理時間は約 0.6fps であった．

図 2 上段が入力画像，中段が基準画像（19 フレーム目）に対して入力画像を射影変換した結果，下段が射影変換画像列に時空間画像フィルタを適用し前景物除去を行った結果である．19 フレーム目より前の結果画像はそのフレームの時点での結果を表示している．異なる位置・角度から情報を蓄積していくことで除去結果が改善されていくのがわかる．また，前景物のない状態で撮影した画像が図 4 であり，これと除去結果を比較したところ，ピクセル値の二乗誤差平均が 11.52，PSNR 比が 86.384 であった．

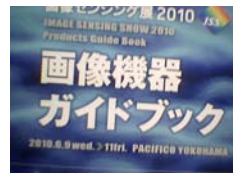


Fig.4 比較用画像

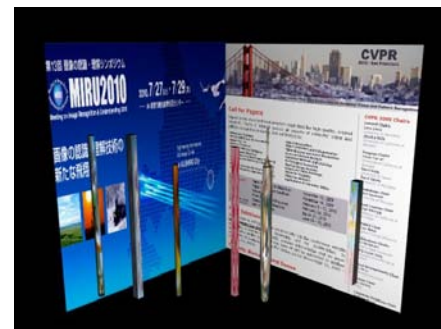


Fig.5 仮想空間実験：入力シーン

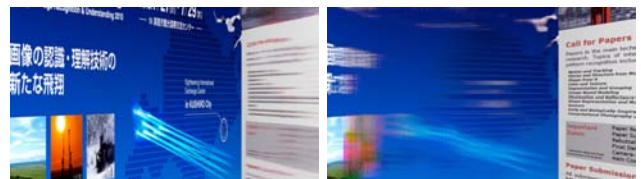


Fig.6 仮想空間実験：左右それぞれの平面での除去結果



Fig.7 仮想空間実験：除去結果を統合した結果

### 5.2 仮想空間によるオフライン実験

背景が二つの平面とそれより手前にいくつかの前景物を置いた構成による仮想空間において実験を行った（図 5）．図 3 上段が入力画像，中段が右の平面について基準画像（0 フレーム目）に入力画像を射影変換した結果，下段が同様に左の平面についての射影変換した結果である．それぞれの射影変換結果に時空間画像フィルタを適用し前景物除去した画像が図 6 である．除去結果を統合した画像が図 7 であり，入力画像から前景物が除去できたことが確認できる．

### 5.3 手持ちカメラによる複数平面実験

くの字形に並べた本を背景，ペンを街路樹や歩行者のような縦長の物体と見立てて，手でカメラを持ち車と同様にスライドするように撮影し実験を行った．図 8 上段が入力画像，中段が右の平面について基準画像（7 フレーム目）に入力画像を射影変換した結果，下段

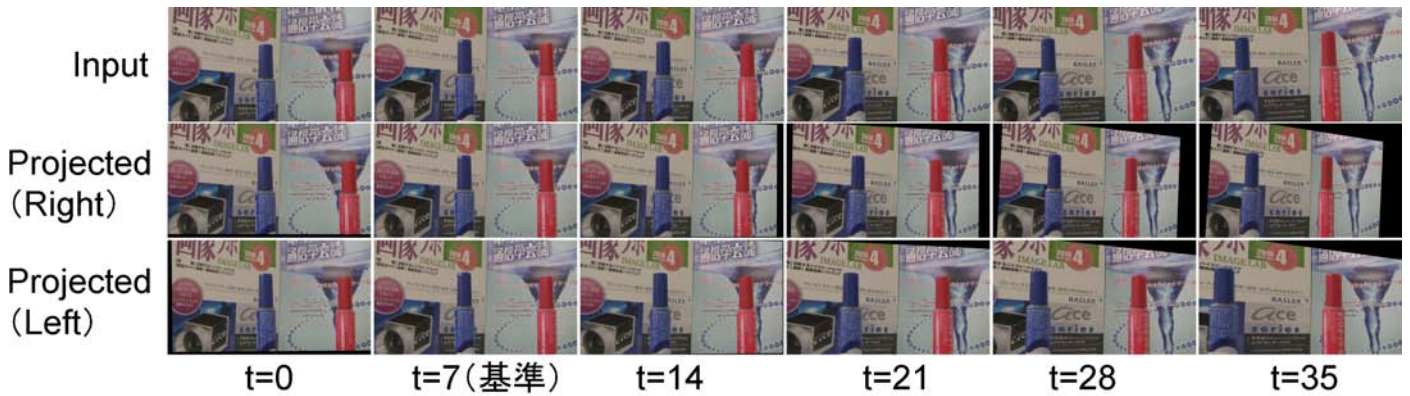


Fig.8 手持ちカメラによる複数平面実験結果

が同様に左の平面についての射影変換した結果である。それぞれの射影変換結果に時空間画像フィルタを適用し前景物除去した画像が図9である。それぞれの平面において入力画像から前景物が除去できたことが確認できるが、右平面においてはエッジ近辺において前景物が除去しきれなかった箇所がある。これはエッジ近辺の画素は同じピクセル値で安定的に現れないこと、前景物のペンの色が一樣であることの両方が重なり合ったためである。除去結果を統合した画像が図10である。



Fig.9 複数平面実験：左右それぞれの平面での除去結果



Fig.10 複数平面実験：除去結果

#### 5.4 車載カメラによるオフライン実験

カメラを車載し、走行しながら撮影した。図11上段が入力画像、下段が背景のビル群を平面として基準画像(20フレーム目)に入力画像を射影変換した結果である。射影変換結果に時空間画像フィルタを適用し前景物除去した画像が図12である。正しく前景物の街路樹や高架橋が除去されていることが確認できるが、右

端部分の街路樹が除去しきれしていないのは背景が一度も見えていないためである。

#### 6. 考察

提案手法では、背景・前景物・カメラの位置関係が重要である。本章ではこの位置関係について述べる。

提案手法と背景についてまとめる。失敗する可能性が高いシーンとは、ベースラインが短かく前景物の遮蔽により背景がほとんど現れない場合、背景平面の特徴が少なく平面検出ができない場合、背景が球面形状を扱う場合などが考えられる。しかし、提案オフライン手法では車を移動させながら撮影するため十分なベースラインを得ることができる。さらに、背景のビル群には看板や窓枠など特徴となるものが豊富に存在し、特徴がほとんど得られない場合は複数平面となることはなく、一枚平面として扱うことができる。また、球面形状のビルが存在する割合は低いといえる。よって提案手法が想定するシーンでは有効に働くことが期待できる。背景が全く見えないシーンについては本論文の対象外であり、inpaintingなどを活用した方がよい。

次に提案手法と前景物についてまとめる。

#### 提案手法が効果的に働く前景物の条件

1. 複数個の場合、個数に関わらず一括に除去できる。
2. 縦長の物体である場合、車載カメラの移動方向と垂直になり、遮蔽をずらしながら撮影できる。
3. 背景よりもカメラに近く存在する場合、遮蔽のずらしが容易にできる。

#### 提案手法が効果的に働かない前景物の条件

1. 背景近くに存在する場合
2. カメラと同じ方向にカメラに近い速度で移動する場合、両方の条件共に背景が見えにくい場合であり、除去が失敗しやすい。

また、カメラに関しては広角なものほど、異なる位置から撮影したときに前景物の裏に存在する背景が映る可能性が高くなるため提案手法に適している。

#### 7. まとめ

本論文では、時空間画像解析を利用した背景が複数の平面の場合のオフライン処理と、背景が一枚平面の場合のオンライン処理の前景障害物除去に関して述べ



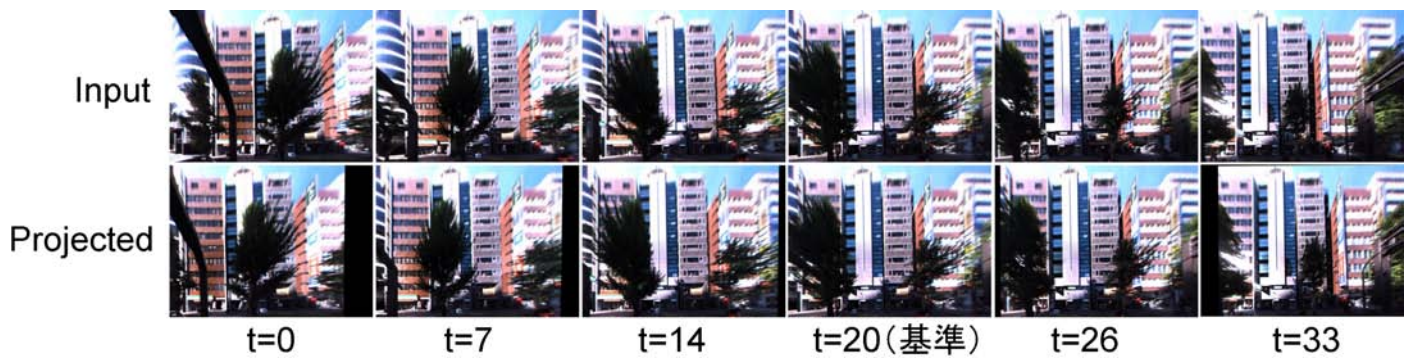


Fig.11 車載カメラによる実験結果



Fig.12 車載カメラによる実験：除去結果

た．オフライン処理では，背景の平面を推定し，各平面ごとに時空間画像フィルタによる前景障害物除去を行い，これらを統合することでシーン全体の前景障害物を除去する手法を提案した．オンライン処理では，オフライン処理の方法を元に処理時間の短縮を除去可能条件を限定することで実現した．最後に，USBカメラ，仮想空間・屋外で撮影したデータを用いて実験を行ったところ，網などの複雑な形状や，背景が自由に位置された平面空間において正しく前景障害物を除去できることを確認した．

#### 参 考 文 献

- 1) Yin Li, Jian Sun, Chi-Keung Tang and Heung-Yeung Shum, "Lazy Snapping", SIGGRAPH 2004, Vol. 23, pp. 303-308.
- 2) C. Rasmussen, T. Korah, "Spatiotemporal inpainting for recovering texture maps of partially occluded building facades" In Proceedings of the IEEE ICIP. 2005.
- 3) Yasuyuki Matsushita, Ko Nishino, Katsushi Ikeuchi, Masao Sakauchi, "Illumination Normalization with Time-Dependent Intrinsic Images for Video Surveillance", IEEE Trans. Pattern Analysis and Machine Intelligence., 26(10): 1336-1347 (2004).
- 4) Takaaki Shiratori, Yasuyuki Matsushita, Sing Bing Kang, and Xiaoou Tang, "Video completion by motion field transfer", In Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR), volume 1, pages 411-418, June 2006.
- 5) 羽下哲司, 鷲見和彦, 橋本学, 関明伸, 黒田伸一, "時空間画像処理による注目距離対象の強調に基づいた遠隔監視システム", 電子情報通信学会技術報告 (PRMU97-126), pp.41-46 (1997).
- 6) 福地 功, 山下 淳, 金子 透, 三浦 憲二郎, "カメラの回転を用いた時空間画像からの水滴除去", 第13回画像センシングシンポジウム講演論文集, 横浜, June 2007.
- 7) Rahul Garg, Hao Du, Steven M. Seitz and Noah Snavely, "The Dimensionality of Scene Appearance", In Proceedings of the IEEE International Conference on Computer Vision (ICCV), Kyoto, 2009.
- 8) D. Lowe, "Distinctive Image Features from Scale-Invariant Keypoints", Int'l J. Computer Vision, vol. 2, no. 60, pp. 91-110, 2004.
- 9) J. M. Morel and G. Yu, ASIFT: A new framework for fully affine invariant image comparison, SIAM Journal on Imaging Sciences, vol. 2, no. 2, pp. 438-469, 2009.
- 10) Herbert Bay, Andreas Ess, Tinne Tuytelaars, Luc Van Gool, "SURF: Speeded Up Robust Features", Computer Vision and Image Understanding (CVIU), Vol. 110, No. 3, pp. 346-359, 2008.
- 11) 山口拓真, 川崎洋, 古川亮, 中山利宏, "画素単位のレジストレーションによる3次元形状物体画像の超解像処理", 第12回画像の認識・理解シンポジウム, 2009.
- 12) J. Astola, P. Haavisto, and Y. Neuvo, "Vector Median Filters", Proc. of the IEEE, Vol.78, No.4, pp.678-689, Apr. 1990.