個人視点映像を一覧するための広視野貼り合わせ画像群の自動生成

貼り合わせの良さに基づいた画像の選択とグループ化

松井　研太†　　近藤　一晃†　　小泉　敬寛†　　中村　裕一†

† 京都大学　〒606–8501 京都市左京区吉田本町

E-mail: †{matsui,kondo,koizumi}@ccm.media.kyoto-u.ac.jp, ††yuichi@media.kyoto-u.ac.jp

あらまし　個人視点映像から複数の広視野貼り合わせ画像を自動生成する方法について述べる．我々は貼り合わせに適した画像を選択するための指標を既に提案している．本稿ではそれに加えて計算量の削減と映像分割の方法を合わせた自動生成の手法を提案する．計算量の削減では局所的な少数画像の貼り合わせの良さに基づいて近似的に枝刈する．映像分割では良い貼り合わせ画像が得られるような分割を先頭から順に行う．計算量を削減することで良い貼り合わせとなる組が網羅できなくなるものの，実用上は問題ない質の貼り合わせ画像が得られることが確認された．広視野貼り合わせ画像の自動生成も併せて行い，個人視点映像を一覧するような画像列が得られることを確認した．

キーワード　個人視点映像，広視野貼り合わせ画像，映像要約

Automatic Synthesis of Stitched Wide FOV Images for Reviewing FPV Videos

Selecting and Grouping Images Based on Stitching Criteria

Kenta MATSUI†, Kazuaki KONDO†, Takahiro KOIZUMI†, and Yuichi NAKAMURA†

† Kyoto University　Yoshidahonmachi, Sakyo-ku, Kyoto-shi, Kyoto, 606–8501 Japan

E-mail: †{matsui,kondo,koizumi}@ccm.media.kyoto-u.ac.jp, ††yuichi@media.kyoto-u.ac.jp

Abstract　This paper reports a technique to automatically synthesize stitched wide FOV images from a first person view (FPV) video for reviewing it. For this purpose, we had proposed criteria to select suitable images for stitching. In this paper, we solve remained issues ; calculation cost and temporal segmentation for fully automatic synthesis. The proposed method approximately reduces the number of image combination (hypotheses) that must be evaluated for the image selection, based on pixel consistency among a small number of local images. The input video is temporally segmented by the image combination with highest stitching score, sequentially. Through the experiments to analyze degree of reduction and how many image combination with high score are dropped, we have confirmed that we can get enough well stitched wide FOV images from reduced hypotheses.

Key words　First person view video, Wide FOV images, video summarization

1. はじめに

頭部に装着されたカメラによって撮影された映像は個人視点映像と呼ばれている．個人視点映像はカメラ装着者の体験を一人称視点で記録できるため，映像メディアとしてのライフログ [1], [2] や体験活動の記録 [3], [4] として幅広く用いられている．しかし，体験を記録した個人視点映像には，映像の質が悪い，長時間に及ぶため閲覧する際の負担が大きいという問題がある．本研究では後者の負担を軽減することを目的としている．個人視点映像に限らず，映像をそのまま閲覧するには記録した時間と同程度かそれ以上の時間が必要とされている．集団で行う体験活動では数時間 × 参加者人数分，ライフログでは日常生活と同じだけの長さの個人視点記録となるため，閲覧には多大な労力を要する．場合によっては現実的な時間に収まらない．

長時間の映像を効率的に閲覧するための試みとして，尺の全体を複数枚の静止画で一覧表示する方法がある [5]〜[9]．このような一覧表示は，映像のおおよその内容を把握できる，映像記録を再生して視聴したい場所を簡単に探すことができる，などの利点を持つ．我々はこの方法を個人視点映像に適用して効率的な閲覧を目指している．静止画列による長時間映像の一覧

表示に関する従来手法では，映像からどのフレームを取り出して用いるか，が主な課題とされてきた．しかし，映像から抜き出した一枚の画像から対象のシーン・状況を把握することは難しい．特に個人視点映像は状況を把握しやすいように撮影・編集されてはいないため，この問題はより深刻となる．本研究では，個人視点映像に含まれている画像を貼り合わせることで広視野画像を作成し，それらを用いて映像全体の一覧表示に用いる仕組みについて検討を進めてきた (図 1)．

このような背景から，我々は良い貼り合わせ画像を得るための画像選択手法についての提案 [10], [11] を過去に行っている．一般に，個人視点映像を構成している全ての画像，あるいは無作為に選んだ画像を用いても，良い貼り合わせ画像を得られるとは限らない．良い貼り合わせ画像を得るためには，入力画像から貼り合わせに適した画像の組み合わせ（以後，部分画像と呼ぶ）を選択する必要がある．文献 [10], [11] で提案した手法は，貼り合わせに用いる部分画像とそれらの位置合わせ情報から，貼り合わせた画像の良さを導出するもので，その値に基づいて良い貼り合わせとなるような部分画像を探すことができる．ただし，一覧表示を作成するにはこの提案だけでは不十分であり，部分画像の数に起因する計算量の増大と画像のグループ化の大きく分けて２つの課題が残されている．本稿では以下のようなアプローチでそれらを解決し，個人視点映像を一覧表示するための広視野貼り合わせ画像を自動生成する手法を提案する．

- 計算対象となる部分画像数の削減

最も良い貼り合わせとなるような部分画像を探すには，考えられるすべての部分画像について上記の指標値を計算する必要がある．ところが，入力画像の枚数が増大するにつれ組み合わせの数が膨大になってしまい，処理が現実的な時間で終わらない．原理的に言えば，入力画像列のそれぞれについて貼り合わせに用いるかどうかの 2 通りが考えられるため，N 枚の入力画像列に対する部分画像の総数は $O(2^N)$ である．数時間の個人視点映像だと N は十数万枚などになる．これでは組み合わせ爆発が起こることは容易に想像できる．貼り合わせは互いに重畳領域を持つ場合にのみ可能であるため，重畳領域を持つものに限った組合せを扱うことで部分画像の数を削減することはできる．しかし，そうして絞ったとしてもまだまだ膨大な数の組合せとなる．

そこで本稿では，以下に挙げるような局所的な少数の画像の貼り合わせの良さに基づいた手法を提案する．

（1） 時間近傍にある二枚の画像間の貼り合わせの良さから，映像内容の変化量を求め，ほとんど変化していない画像列は一枚の代表画像で置き換える．

（2） うまく貼り合わせられない少数の画像の組み合わせを事前に列挙しておき，その組み合わせを含む部分画像を除外する．

前者は入力画像そのものの枚数を減らす方法，後者は組み合わせの数を減らす方法となっており，これらを用いて計算対象とする部分画像の数を削減する．

- 映像分割

広視野貼り合わせ画像により個人視点映像を一覧表示するには，どの区間を一枚の貼り合わせ画像で表すか，すなわち画像のグループ化もしくは映像の分割を決める必要がある．入力映像を一定時間ごとに区切ることが最も単純な分割方法であるが，映像の内容を考慮していないので良い一覧表示にはなりづらい．映像内容に基づいた分割・一覧表示には，ショットの切り替わり [6] やシーンの切り替わり [5] を検出する方法などが主流である．これらの手法では，まず映像の分割を行い，次に分割された短時間映像の各々から代表画像を一枚つづ取り出して一覧表示を構成する．これに対し本提案では良い広視貼り合わせ画像が生成できるように映像の分割を行う．言い換えれば，従来手法が意味に基づいて，提案手法は二次元画像上での幾何学的整合性に基づいて映像を分割するとみなすこともでき，これらはうまく組み合わせることでより効果的な一覧表示画像を生成できると考える．

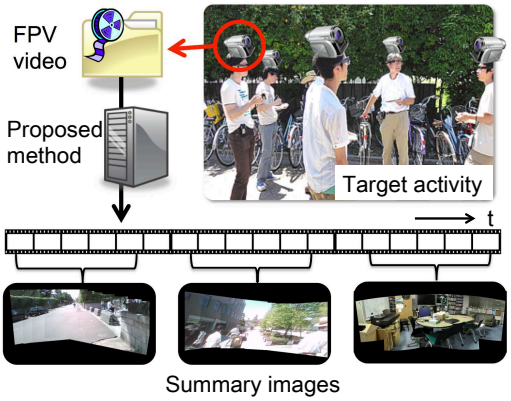


図 1　個人視点映像を一覧する広視野貼り合わせ画像の概要

2. 貼り合わせの良さを用いた画像選択

本稿で提案する部分画像数の削減と映像分割の手法に大きく関わるため，まず，我々が過去に提案した貼り合わせの良さの導出法について述べる．詳細については文献 [10], [11] を参照されたい．

2.1 画像選択の必要性

複数の画像を貼り合わせて一枚の広視野画像を生成する手法は既にいくつか提案されている [12]〜[15]．多くの貼り合わせ手法では，カメラを水平方向に回転させながら撮影することを前提としており，入力画像を事前に想定した幾何学変換に基づいて貼り合わせている．例えば [14], [15] では，ホモグラフィ変換による位置合わせ方法，およびそれで近似できなかった微小なズレを補正する方法が提案されている．しかし，個人視点映像を対象として貼り合わせを行うには，それに加えて貼り合わせに適した画像を選択することが求められる．体験記録としての個人視点映像ではあるがままの様子を収めることが重要である．故にカメラ装着者が一般のシーンで自由に行動することを前提としている．このため，撮影された画像全てが貼り合わせに適しているとは限らない．例えば，振り向きや見渡しといった行動が発生したときには，モーションブラーやローリングシャッター効果などを含んだ質の悪い画像が記録されてしまう．また，カメラ移動などにより視差が発生した画像同士はぴったりと重

ね合わせることが難しい．視野内に動物体が含まれていた場合も同様である．このような不適な画像や画像の組み合わせを除外するような仕組みが必要となる．

2.2 貼り合わせの良さの導出法

上記問題を解決するために，貼り合わせの良さの程度を数値化し，それに基づいて画像選択を行う．この「貼り合わせの良さ」を見た目の良さとして考えれば，対象シーンにおける時空間的な幾何は必ずしも保存される必要はなく，貼り合わせた結果「複数枚の画像が矛盾なく重なり合っていること」「広視野となっていること」が重要である．これら二つの指標を，同一の尺度で，かつ任意の枚数の重ね合わせに対しても同様に扱える仕組みを提案した．貼り合わせにより重なり合った複数画素値がシーン中の同一点を観測したものかどうかを確率的に調べることで，画像が矛盾無く重なっているかどうかを検証するものである (図 2)．具体的には，複数画素値を観測とみなしたときのシーン中の画素値の事後分布が特定の値付近に集中している度合いを貼り合わせの良さとして出力する．この枠組みを式で記述すると以下のようになる．

$$Q(\{y_t\}) = F(p(x|\{y_t\})) \tag{1}$$

$$p(x|\{y_t\}) = \frac{p(x)p(\{y_t\}|x)}{p(\{y_t\})} \tag{2}$$

Q は貼り合わせの良さ，x はシーン中の真の画像値，$\{y_t\}$ は観測画像上での画素値，$F(\cdot)$ は分布の集中度を出力する関数を示す．ただし画像 $\{y_t\}$ は互いに位置合わせ済みとする．このように定式化することで，$\{y_t\}$ の要素数すなわち観測画像の枚数に依らない表現となっている．貼り合わせ画像の広がりは「無観測であった領域が一回の観測を得た」と解釈することで画像の重ね合わせと同一の枠組みで扱うことができる．シーンから画像上の画素値への観測過程は，領域毎に画像の重なり枚数が異なることを考慮して以下のようにモデル化した．

$$y = \alpha(x + n(\sigma_1)) + (1-\alpha)u \tag{3}$$

ここで x, y はシーン中の真の画素値および観測画像における画素値，$n(\sigma_1)$ は雑音ノイズを示すガウス分布，u は一様分布，$0 \leqq \alpha \leqq 1$ は位置合わせのずれやぼけがない確率である．この観測モデルは，位置ずれ・ぼけがない場合には画素値はガウス分布に従った揺らぎを持ち，それらがある場合にはどの値も等しくとりうる，というものである． 位置 i に対応するシーン中の真の画素値 $x(i)$ の事後確率は，複数の観測画素値 $\{y_t(i)\}$ が互いに独立であるという仮定の下で

$$p(x|\{y_t\}) = [p(x(0)|\{y_t(0)\}) \cdots p(x(S)|\{y_t(S)\})]^T \tag{4}$$

$$p(x(i)|\{y_t(i)\}) = p(x(i)) \prod_{t=1}^{T} h(x(i), y_t(i))$$
$$h(x(i), y_t(i)) = \begin{cases} 1.0 & y_t(i) = \phi \\ \frac{p(y_t(i)|x(i))}{p(y_t(i))} & otherwise \end{cases} \tag{5}$$

のように表される．S は貼り合わせ領域の大きさ，T は入力画像の枚数，$y_t(i) = \phi$ は「位置 i は画像 t の視野外である」ことを示す． 式中の事前分布には一様分布 $p(x(i)) = \frac{1}{X}$ を与え，尤度には (3) 式で示した観測過程より $p(y_t(i)|x(i)) = \alpha G(x(i) - y_t(i), \sigma_1) + (1-\alpha)\frac{1}{X}$ を与える．ここで $G(\cdot)$ はガウス分布，X は画素値の取りうる範囲を示す．

こうして得られた事後分布がどれだけ特定の値付近に集中しているかを $F(\cdot)$ を用いて定量化する (図 3)．まず，特定の値 x_c をシーン中の真の画素値とし最尤推定で決定する．次に，x_c を平均としたガウス分布と事後分布との類似度を計算して集中度とする．複数回観測された位置では複数画素値の整合性に応じた値が，一枚の画像でのみ占められている位置では一回観測の事後分布の形で決まる定数値が算出される．この手続きを式を用いて記述すると以下のようになる．

$$q(\{y_t\}) = \int_X p(x|\{y_t\}) G(x - x_c, \sigma_2) dx \tag{6}$$

$$x_c = argmax\, p(x|\{y_t\}) \tag{7}$$

シーン中の真の画素値が観測過程以外の要因，例えば光源変化や撮影方向によるアピアランスの変化，による画素値の揺らぎを分散 σ_2 のガウス分布でモデル化しており，曖昧さを含めた上での尤度を計算していることに相当する．(6) 式は各画素について独立に計算されるため，貼り合わせ領域 S について積算することで最終的な貼り合わせの良さ Q を得る．

$$\begin{aligned} Q(\{y_t\}) &= F(p(x|\{y_t\})) \\ &= \textstyle\sum_i^S f(p(x(i)|\{y_t(i)\})) \\ &= \textstyle\sum_i^S \log \int_X p(x(i)|\{y_t(i)\}) G(x(i) - x_c(i), \sigma_2) dx \end{aligned} \tag{8}$$

ここで対数をとって積算しているのは上記集中度が尤度として計算されているからである．

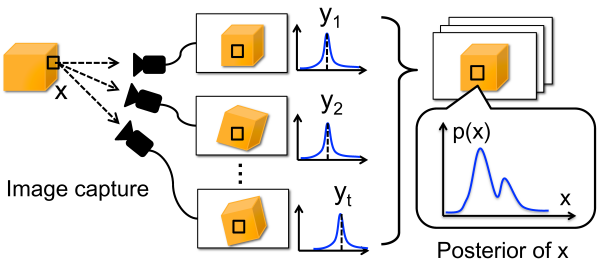
貼り合わせの良さを導出する一連の手続きで用いられているパラメータは $\sigma_1, \sigma_2, \alpha$ の 3 つである．これらは重ね合わせの良さと広がりの良さをバランスを制御するパラメータとして働く．例えば，σ_1, σ_2 を大きく設定すると，より灘らかな事後分布が (5) 式で得られやすく，また (6) 式で高スコアを得やすくなる．結果的に複数画素値の整合性が低くても広い貼り合わせ画像となるような部分画像が高い値をとるようになる．

この手法は複数枚の低解像度画像を合成して高解像な画像を生成する超解像の手法，特に観測モデルと複数回観測による事後分布推定はベイズ超解像 [16]〜[19] と似た考え方である．画像の貼り合わせでは，重なりだけでなく広がりも考慮しなければならないこと，シーンの真の画素値を推定するだけでなくその過程で得られる事後確率の分布を利用していること，が大きく異なる点となっている．

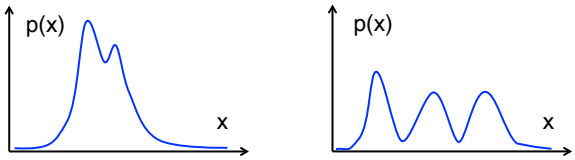
3. 部分画像数の削減と映像分割

3.1 代表画像による類似画像列の置き換え

事後分布 $p(x|\{y_t\})$ が非線形なため (7) 式の 画像の貼り合わせでは，重なりだけでなく広がりも考慮しなければならないこと，シーンの真の画素値を推定するだけでなくその過程で得られる事後確率の分布を利用していること，が大きく異なる点となっている．推定に手間がかかること，(6) 式中のガウス分布



(a) 複数回撮影された同一点の事後分布導出



(b) 整合性が高いときの事後分布　(c) 整合性が低いときの事後分布

図 2　観測モデルと複数回の観測による事後分布

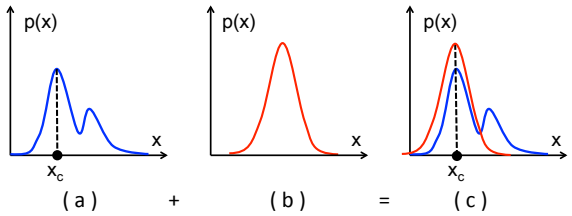


図 3　事後分布の集中度合いの定量化．(a) 推定された事後分布 (b) 参照分布，(c) 畳み込み

の積分項で誤差関数が出現すること，(8) 式に示すように貼り合わせ画像全体について計算する必要があること，などから 1 組の部分画像 $\{y_t\}$ に対する貼り合わせの良さの計算には相応の時間を要する．さらに貼り合わせの良さは用いる部分画像 $\{y_t\}$ 毎に異なるので，良い貼り合わせ画像を網羅的に探すには，考えられるすべての部分画像について計算する必要がある．このような理由から，長時間の個人視点映像をそのままに用いると現実的な時間に計算が収まらない．本節では入力画像の数そのものを減らすことで列挙する部分画像の数を削減する手法について述べる．

個人視点映像の中には，シーンやカメラワークがよく変化する場面もあれば，撮影された内容がほとんど変化しない場面もある．撮影された内容がほとんど変化していない画像列では，その中のどれを用いても同等の貼り合わせ結果となることが予想される．そこで互いに類似した画像は一枚の代表画像で置き換える処理を適用することで，全画像列を用いた場合と同等の入力画像を少ない枚数で構成する．これは画像内容の変化に応じて動的・仮想的にフレームレートを下げることに相当する．本提案では，以下の条件を満たす連続画像を類似した画像列として扱う．

- 画像列中の任意の二枚の画像が十分な重畳領域を持つ
- それらが互いに矛盾なく重なりあう

類似画像列を検出する手続きは以下のとおりである．

まず入力画像列の全体を対象に 2 枚の画像の組を列挙・位置合わせし，重畳領域の面積と重なりの良さを計算する．位置合わせにおける幾何学モデルにはホモグラフィ変換を用いた．これはシーン中の平面を撮影した画像や無限遠およびカメラが回転運動のみをしたときの画像であれば矛盾無く重なり合う変換である．貼り合わせにおいても同様の幾何学モデルを用いている．ホモグラィ変換の推定には局所特徴量の対応に基づいた RANSAC によるロバスト推定を用いた．推定された変換行列を用いて二枚の画像各々を共通座標系（パノラマ座標系）に投影し，その上で重畳領域の面積と重なりの良さを算出する．ただし時間的に遠い 2 枚の組み合わせは，(i) 重畳領域がない場合が多い，(ii) 重畳領域の有無を誤推定しやすい．(iii) 計算量が増える，ため，時刻差が閾値 ne_{th} フレーム以下の組だけを対象として計算する．また対象の 2 枚が重畳領域を持たないこともある．その場合には明らかに間違った変換行列が得られるため，以下の指標を用いて判別を行う．

(1) 変換後の 2 枚の画像の面積比ホモグラフィィ変換が正しく推定されていたならば変換先のパノラマ座標系において 2 枚の画像の面積は大きく変わらないはずである．そこで以下の条件を満たさない場合は誤推定とみなし，対象の 2 枚に重畳領域はないと判断する．

$$1 - r_f < \frac{S(I_a)}{S(I_b)} < 1 + r_b \tag{9}$$

ここで，I_a, I_b はパノラマ座標系に変換された 2 枚の画像，$S(\cdot)$ は面積を出力するオペレータ，r_e は許容する面積変化の閾値である．

(2) 変換前後における 4 端点の上下左右位置関係の保存一般シーンを撮影した場合，2 枚の画像が左右や上下に反転して重なり合うことはほとんどない．2 枚の画像のどちらかに変換前後で上下左右位置関係が逆転するような画像 4 端点が存在した場合は，同様に対象の 2 枚に重畳領域はないと判断する．

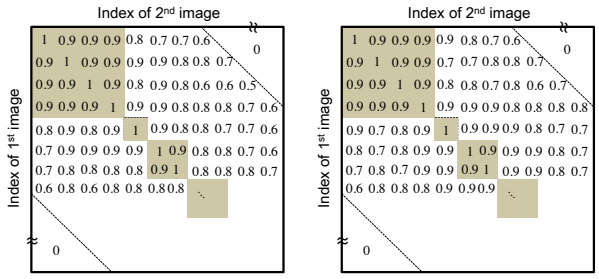
以上の条件を満たした 2 枚の画像に対して，重畳領域の面積 $r_{ij}(i, j$ は画像フレーム番号) を計算する．これは (1) で既に求められている値を利用して

$$r_{ij} = \frac{2S(I_a \cap I_b)}{S(I_a) + S(I_b)} \tag{10}$$

と定義する．重なり合わせの良さ w_{ij} には画素毎における画素値の差の平均を用いる．

$$w_{ij} = \frac{1}{S(I_a \cap I_b)} \sum_{i}^{I_a \cap I_b} \frac{X - |I_a(i) - I_b(i)|}{X} \tag{11}$$

(8) 式で示した貼り合わせの良さを用いることもできるが計算量の問題から単純な (11) 式を採用した．なお，2 枚の重ね合わせに関しては (11) 式でも十分に良く働くことが予備実験で確認されている．条件 (1)(2) により重畳領域がないと判断された場合は $r_{ij} = w_{ij} = 0$ とする．重畳領域の面積および重ね合わせの良さはどちらも任意の 2 枚の画像間で定義される値なので，行列の形 $R = \{r_{ij}\}, W = \{w_{ij}\}$ で表すことができる．順序は無関係なので $r_{ij} = r_{ji}$, $w_{ij} = w_{ji}$ である．行列表現した重畳領域の面積 R，重ね合わせの良さ W の例を図 4 に示す．行列 R, W の対角成分 r_{ii}, w_{ii} は同一の画像に関する重畳領域面積お



(a) 重畳面積 R　　(b) 重ね合わせの良さ W

図 4　画像間の関係の行列表現

よび重ね合わせの良さを示しているので必ず 1 となる．

そこで対角成分を中心としており全ての要素が閾値以上の部分正方行列を類似画像列として検出する．先頭を I_s とする類似画像列の末尾 I_e は

$$e = max(e_k)\, s.t. \sum_{i=s}^{e_k}\sum_{j=s}^{e_k} f(r_{ij}, w_{ij}) = (e_k - s + 1)^2 \quad (12)$$

$$f(r_{ij}, w_{ij}) = \begin{cases} 1 & r_{ij} > r_{th}\,,\, w_{ij} > w_{th} \\ 0 & otherwise \end{cases} \quad (13)$$

のようにして決定される．r_{th}, w_{th} は画像が類似しているとみなす閾値を示す．このようにして得られた類似画像列 $[I_s, I_e]$ を時間中央の代表画像 $I_c, c = int((e-s)/2)$ で置き換える．$int(\cdot)$ は整数化オペレータである．

上記のような類似画像列の検出と代表画像による置き換えで，どの程度画像枚数の削減されるかは映像の内容変化に依存する．カメラがほとんど動いておらずシーンも静的であるような区間は代表画像が少数となるため削減率が高く，カメラが動いていたりシーンが動的である場合にはあまり削減されない．加えて本手法はどの程度まで類似した画像を 1 枚とみなすかという近似でもある．近似の度合いは部分行列を取り出す際の閾値 r_{th}, w_{th} によって変化する．削減率と近似による弊害（重要な部分画像がどの程度破棄されてしまうのか）の関係については実験の章で検証を行う．

3.2　整合性の低い 3 つ組画像の同時不使用

良い貼り合わせ画像が満たすべき条件として，(8) 式で導出される画像全体で見た貼り合わせの良さに加えて，局所的に見たときの重ね合わせの良さも重要である．たとえ多くの領域で複数画像がぴったりと重なり合っており，全体として広視野を有していたとしても，特定の領域で大きな不整合があってはならない，という条件をつけることは自然と考える．この条件をうまく利用して部分画像の列挙数削減も同時に行うことを試みる．

特定の領域での不整合は局所的な画像群の重ね合わせの良さで表すことができる．ここで「局所的な少数の画像 $\{I_{local}\}$ の重畳領域において重ね合わせが悪い場合には，新たに画像を加えてもやはり悪い重ね合わせとなる」という仮定をおく．すると $\{I_{local}\}$ を含むような部分画像は，画像全体での貼り合わせの良さを計算するまでもなく悪い貼り合わせであるとみなして近似的に列挙数の削減を行うことができる．例えば 3 枚の画像がうまく重ね合わせられないとき，さらに 4 枚目・5 枚目を追加しても重ね合わせが大きく改善されることはない，といった具合である．この方法は原理的に言えば任意の少数枚の画像の組み合わせについて適用できる．しかし，あまりに少ない枚数の組では他の画像が追加されたときに重ね合わせが十分に改善されてしまう可能性が高い．多数の画像の組を対象とすればその可能性は低くなるが，そのような組の数は少ないので削減率が低くなってしまう．そこで本稿では 3 枚の画像の組を用いて上記削減を行う．

まず共通した重畳領域をもつ 3 枚の画像の組み合わせを列挙する．考えうる 3 つ組画像は $O((N-ne_{th})\cdot ne_{th}C_3)$ 組存在するので全てについて検証していては計算量が増大してしまうため，3-1 節で導出されている行列 R を活用して効率的に列挙する．R の各要素は任意の 2 枚が重畳領域を持っているかどうかを示しているので，その情報からまず 2 枚同士が互いに重畳領域をもつような 3 つ組画像 $I_{abc} = (I_a, I_b, I_c), s.t.\quad r_{ab} > 0, r_{bc} > 0, r_{ca} > 0$ を列挙する．

この条件を満たす I_{abc} でも 3 枚全てに共通する重畳領域をもつとは限らない．また重畳領域が狭い場合，全体の貼り合わせの良さに対する影響は少ないので除外したい．そこで閾値 ts_{th} 以下の重畳領域しか持たないような組はここで破棄する．こうして列挙された 3 枚の画像の組について重畳領域における重ね合わせの良さ $Q_3(I_{abc})$ を計算する．Q_3 は I_{abc} を入力画像としたときの貼り合わせの良さを重畳領域のみで算出し，その面積 S_{ov} で正規化することで得る．

$$Q_3(I_{abc}) = \frac{\sum_i^{S_{ov}} f(p(x(i)|I_a(i), I_b(i), I_c(i)))}{S_{ov}} \quad (14)$$

$Q_3(I_{abc})$ が閾値 co_{th} 以下であったとき I_{abc} は同時には使用してはいけない 3 つ組画像として登録し，部分画像を列挙する際にフィルタとして用いる．

この手法も代表画像による類似画像列の置き換えと同様に近似的な枝刈りであることに注意したい．仮定が満たされない，すなわち，追加の画像によっては重ね合わせが大きく改善されてしまう場合も存在する．部分画像数の削減率と網羅性への弊害はトレードオフの関係にあり閾値 co_{th} によって制御される．この特性については実験の章で検証する．

3.3　貼り合わせ結果に基づいた映像分割

入力映像を事前に分割して各々から一覧表示用の画像を生成するのではなく，良い貼り合わせとなるように映像を分割することを考えよう．貼り合わせの良さの値が大きい部分画像を 1 組選び出し，続いてそれに含まれている画像を使用していない部分画像から貼り合わせの良いものを再度選ぶ，というように順に広視野貼り合わせ画像の生成と映像の分割を順に行う方法が考えられる．しかしこの方法は考えうる全ての部分画像に関して貼り合わせの良さを計算しておく必要があるため，ここでもやはり計算量の問題が発生する．なぜなら，3-1,3-2 節で述べた提案手法は確かに部分画像数を削減するが，母数すなわち大本の入力画像の枚数が多いとやはり削減しきれずに組み合わせ爆発が起こるからである．

そこで本稿では，一枚の広視野貼り合わせ画像で表現してよ

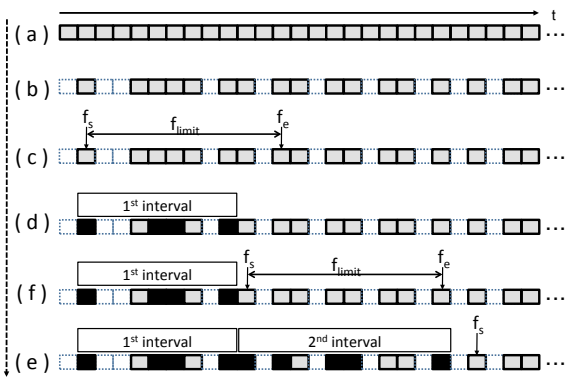


図 5　広視野画像の生成に基づく映像分割

い時間長 f_{limit} を設けることで一度に入力できる枚数を制限し，さらに入力映像の先頭から順に貼り合わせ画像の生成と分割を行う手法をとる．これにより列挙される部分画像の数に上限が設定されるので計算量を制御することが可能となる．時間的に遠く離れた画像を一枚に収めるてしまうとあたかもそれらが同時に存在していたかのように見えてしまうため，時間長に制限を設けることは広視野貼り合わせ画像で状況を表現する上でも利点があると考える．上記映像分割の方法に加え，部分画像数の削減や貼り合わせの良さの算出も含めた一連の処理の流れを以下に示す．(1) 個人視点映像を連番画像に分解し N 枚の入力画像 $\{I_t\}$ を得る (図 5(a))．(2) 3-1 節で述べた代表画像による類似画像列の置き換えを行う．これにより入力画像が N 枚から N_{rep} 枚の代表画像 $\{I_t^{rep}\} \subset \{I_t\}$ に削減される (図 5(b))．以降の処理はこの代表画像列に対して適用される．(3) 貼り合わせを開始する画像番号 $f_s = 0$ を設定する (5(c))．(4) 貼り合わせに用いてよい代表画像列 $[I_{f_s}^{rep}, I_{f_e}^{rep}], fe = max(f)\, s.t.\, f < f_s + f_{limit}$ を取り出す (図 5(c))．(5) $[I_{f_s}^{rep}, I_{f_e}^{rep}]$ を入力画像として最も良い貼り合わせ画像とそれを構成する部分画像を選ぶ (図 5(d))．(i) 3-2 節で述べた方法で同時に使用してはいけない 3 つ組画像 $\{I_{abc}\}$ を列挙する．(ii) $[I_{f_s}^{rep}, I_{f_e}^{rep}]$ に対して 1 枚の貼り合わせ画像を構成でき，かつ $I_{f_s}^{rep}$ を含むような部分画像 $\{I_p\}$ を列挙する．このとき $\{I_{abc}\}$ を含むような部分画像は除外する．(iii) $\{I_p\}$ の要素 I_p 各々に対して貼り合わせの良さ $Q(I_p)$ を計算する．(iv) 貼り合わせの良さが最大であるような部分画像 $I_p^{best} = argmax\, Q(I_p)$ を選択し，広視野貼り合わせ画像を生成する．画像の合成は単純に各画素における複数画像の画素値を平均することで行う．(6) $f_s = f_{latest} + 1$ で更新する (図 5(e))．f_{latest} は I_p に含まれている代表画像の中で最も時間的に新しいものの番号を示す．(7) (4)-(6) を代表画像列の末尾まで繰り返す (図 5(e),(f))．

4. 検 証 実 験

4.1　代表画像による類似画像の置き換え

代表画像による類似画像列の置き換えは近似的な削減方法であることは既に述べた．本実験では計算量に対する削減の効果と削減したことによる部分画像の網羅性の変化について検証を行う．

実験の手順は以下のとおりである．

(1) 個人視点映像から連番の入力画像 $\{I_t\}$ を取り出す．(2) 類似画像列を置き換える代表画像 $\{I_t^{rep}\} \subset \{I_t\}$ を得る．(3) $\{I_t\}, \{I_t^{rep}\}$ のそれぞれに対して 1 枚の貼り合わせ画像を構成できるような部分画像 $\{I_p\}, \{I_p^{rep}\}$ を列挙する．(4) $\{I_p\}, \{I_p^{rep}\}$ の要素各々に対して貼り合わせの良さを計算する．代表画像による置き換えの効果のみを検証するため，同時に使用してはいけない 3 つ組画像のフィルタは用いない．入力画像は PointGrey 社製の IEEE1394 カメラで撮影した映像から 14 枚の連続画像 (640×480) である．貼り合わせの良さを計算する上でのパラメータは $\sigma_1 = 5, \sigma_2 = 30, \alpha = 0.5$，ホモグラフィ行列の誤推定を判断するパラメータは $r_f = 0.2, r_b = 0.1$ とした．

削減の効果は列挙される部分画像の数をどの程度まで減らすことができたか，すなわち削減率 $r_{rep} = \frac{N(\{I_p^{rep}\})}{N(\{I_p\})}$ で評価する．$N(\cdot)$ は要素数を返すオペレータである．類似画像とみなす閾値 r_{th}, w_{th} を様々に設定したときの削減率を表 1 に示す．なお本実験では $r_{th} = w_{th}$ とした閾値の調整により代表画像の枚数が変化し，その結果，列挙される部分画像の数が大幅に削減されることが確認された．部分画像の数は組み合わせ数大きく依存するため，代表画像が少数減るだけでも大きな削減効果がある．実験では 14 枚の連番画像を用いたが，これ以上の枚数を提案手法なしで扱った場合，組み合わせ爆発により計算が終了しなかった．このことからも連続した類似画像列を代表画像で置き換える効果は大きいと考える．

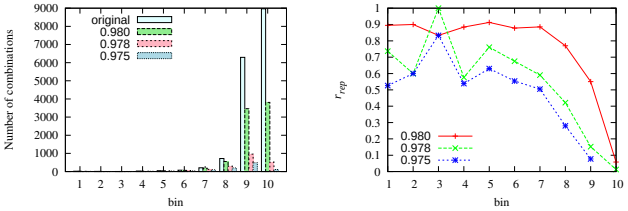
網羅性の変化では，$\{I_p\}, \{I_p^{rep}\}$ で貼り合わせの良さの分布がどのように変わったのかを分析する．計算された貼り合わせの良さの分布を 10 段階のヒストグラムで可視化したもの，および各ビンにおける (=同等の貼り合わせのよさをもつ) 部分画像数の比を図 6 に示す．本研究では貼り合わせの良さの値が大きい部分画像が重要である．すなわち，代表画像による置き換えがあっても，大きなビン番号では部分画像の数が減らないことが求められる．しかし，実験結果はその要求どおりにはならず，むしろ 9 や 10 といった大きなビン番号=重要な部分画像が大きく漏れ落ちている．提案手法の特性から，全体に渡って同程度に削減されることは予想されたが，それとも異なる結果となった．これは複数枚の (類似) 画像を 1 枚で置き換えたことが原因だと考えられる．類似した画像群は重ね合わせれば重ね合わせるほどに (5) 式で定義した分布は先鋭化するので良い評価を得ることになり，相対的に 1 枚しか用いない場合には低い評価値となってしまうわけである．ただし，現状の画像合成方法では複数枚がぴったり重ね合わさった場合と 1 枚しか用いない場合とでは同等の結果となるため，実用上はほぼ問題ないことは確認されている．それぞれの閾値設定で最も良いと評価された広視野貼り合わせ画像を図 7 に示す．用いた画像は異なっているが，人が見てほぼ同様の結果となっていることがわかる．

4.2　整合性の低い 3 つ組画像の同時不使用

3 つ組画像の同時不使用を用いた組み合わせ数の削減についても削減効果と網羅性にトレードオフの関係があるため同様に検証を行った．実験方法は 4-1 節で示した方法とほぼ同じである．

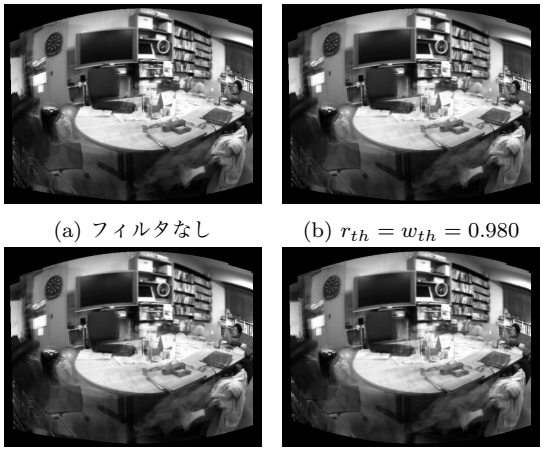
表 1　類似画像の置き換えによる計算量の削減率

閾値 $r_{th} = w_{th}$	フィルタなし	0.980	0.978	0.975
代表画像の枚数	14	13	11	10
列挙された部分画像の数	16366	8175	2035	1010
削減率 r_{rep}	1.00	0.50	0.12	0.06



(a) 頻度の分布　　(b) フィルタ無しとの頻度の比

図 6　類似画像の置き換えによる網羅性の変化



(a) フィルタなし　(b) $r_{th} = w_{th} = 0.980$

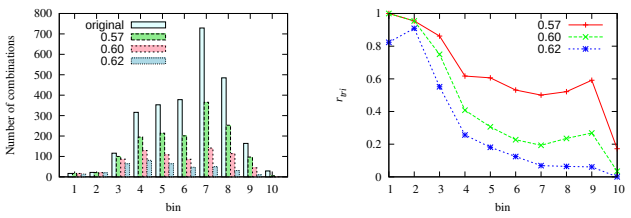
(c) $r_{th} = w_{th} = 0.978$　(d) $r_{th} = w_{th} = 0.975$

図 7　それぞれの閾値で最も良いと評価された広視野貼り合わせ画像

(1) 個人視点映像から連番の入力画像を取り出して代表画像 $\{I_t\}$ で置き換える．(2) 同時に使用してはいけない 3 つ組画像 $\{I_{abc}\}$ を列挙する．(3) $\{I_t\}$ に対して 1 枚の貼り合わせ画像を構成できるような部分画像 $\{I_p\}$，および $\{I_{abc}\}$ を含むような部分画像を除外した部分画像 $\{I_p^{tri}\}$ を列挙する．(4) $\{I_p\}, \{I_p^{tri}\}$ の要素各々に対して貼り合わせの良さを計算する．本実験では同様のカメラで撮影した 50 枚の連続画像を $r_{th} = w_{th} = 0.8$ で代表画像化した 15 枚を入力画像として用いた．他のパラメータについては 4-1 節で示した値と等しくした．削減率 $r_{tri} = \frac{N(\{I_p^{tri}\})}{N(\{I_p\})}$ の結果を表 2 に，貼り合わせの良さの分布を図 8 に示す．同時使用しない 3 つ組画像が増えるにつれて列挙される部分画像の数が削減されているが，表 1 のように指数的に減るわけではなく，3 つ組画像の数と削減率はほぼ線形の関係にある．ただし，この関係を明らかにするためにはさらなる検証実験が必要と考える．4-1 節と同様に貼り合わせの良さの値が高い部分で部分画像の数が保存されている分布が望まれるが，図 8 ではおおよそその逆の傾向になっている．しかし，比較対象の母数には貼り合わせ画像全体で積算したものも含まれている，すなわち局所的に不整合な領域があるような部分画像も数え上げられている．提案手法の効果を正しく検証するには，3 枚の重ね合わせの良さにより 4 枚以上を用いたときの部分画像を効果的に枝刈りできているか，という観点に絞った実験が必要と考えている．

表 2　3 つ組画像の同時不使用による計算量の削減率

閾値 co_{th}	フィルタなし	0.057	0.060	0.062
同時使用しない 3 つ組画像の数	0	3	7	9
列挙された部分画像の数	2609	1468	748	381
削減率 r_{tri}	1.00	0.56	0.29	0.15



(a) 頻度の分布　　(b) フィルタ無しとの頻度の比

図 8　3 つ組画像の同時不使用による網羅性の変化

4.3　広視野貼り合わせ画像の自動生成

提案手法を用いて個人視点映像から広視野貼り合わせ画像を自動生成した場合，実際にどのような画像が出力されるのか，また撮影時のカメラの動き・シーン構造との関係を調査する．自動生成のアルゴリズムは 3-3 節に示したとおりである．入力映像には歩行と見回しを繰り返す行動時の個人視点映像を用いることで日常生活における行動記録を再現した．入力映像の長さは約 1 分の 791 フレームである．同時に使用しない 3 つ組画像の列挙には $co_{th} = 0.06$，映像分割における一枚に収めてよい時間幅 $f_{limit} = 100$ とした，その他，貼り合わせの良さに用いるパラメータなどは前節と同じ値である．その結果，図 9 に示すように 8 枚の広視野貼り合わせ画像が生成された．歩行時には視差が含まれるような画像が記録されやすいため見回し時よりも貼り合わせが困難である．図 9 から歩行時を境に映像分割が行われており，提案手法が正しく働いていることが確認される．3 枚目・4 枚目の貼り合わせ画像は数秒間同じ場所に立ち止まっていた時刻に対応する．本来は 1 枚の貼り合わせ画像で表したい状況であるが，設定した最大時間幅 $f_{limit} = 100$ によって映像分割されたために 2 枚の貼り合わせ画像に分かれている．加速度センサ等の動きを計測するデバイスなどを併用するれば，f_{limite} の動的設定を通してこの問題は解決できると考えている．現状の実装では約 1 分の個人視点映像を処理するのに数時間の計算を要した．実用には計算量や計算時間を短縮するような更なる工夫が必要である．

5.　おわりに

本稿では個人視点映像から複数の広視野貼り合わせ画像を自動生成する方法について述べた．貼り合わせに用いる部分画像の数が爆発する問題に対しては，局所的な少数画像の貼り合わせの良さに基づいて近似的に枝刈する手法を提案した．また従来のように映像分割を事前に行うのではなく，良い貼り合わせ画像が得られるように分割を行う手法も併せて提案した．実験では，計算量の削減率を網羅性への影響について検証した．計

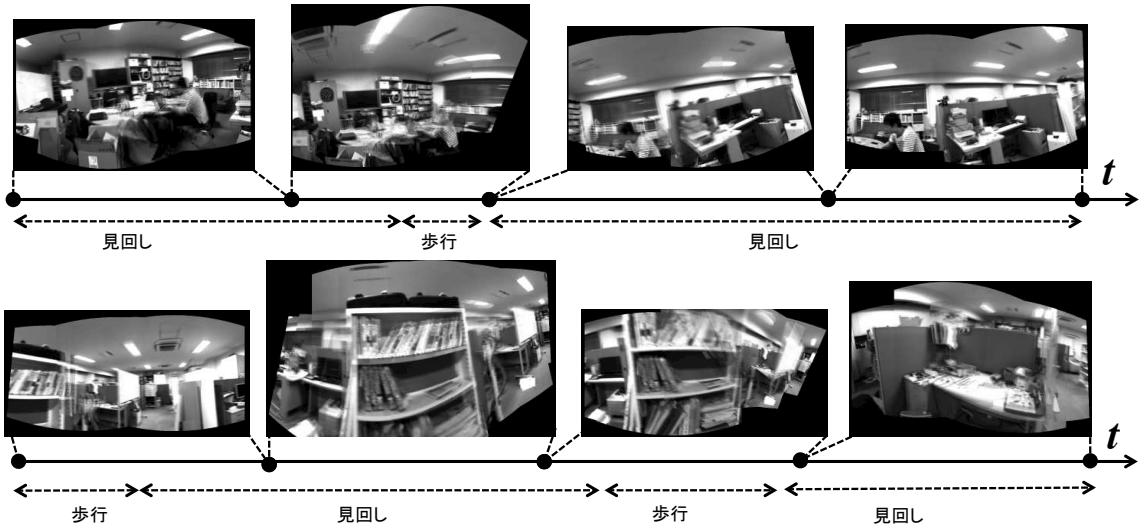


図 9　個人視点映像を入力とした広視野画像貼り合わせ画像の自動生成

算量を削減することで良い貼り合わせとなる組が網羅できなくなるものの，実用上は問題ない質の貼り合わせ画像が得られることが確認された．また広視野貼り合わせ画像の自動生成も併せて行い，個人視点映像を一覧するような画像列が得られることを確認した．

ただし網羅性への影響についても詳細な検証が求められることが明らかとなった．また現状必要とされる計算時間は実用には遠く，計算量削減に関する更なる工夫が必要と考える．加速度センサ等の動き情報から推定されるカメラ運動を追加の情報として活用する手法を現在構想中である．良い一覧表示画像を生成するためには幾何学的な整合性だけでなく，例えば，人物が撮影されていたときには必ず貼り合わせ画像に登場させたい，といったような意味的な側面も考慮する必要がある．カメラ装着者の「体験」をうまく取り出すような仕組みも併せて検討していきたいと考えている．

謝辞　本研究の一部は，科学研究費補助金：集合的個人視点映像を用いた「体験活動を観る・伝える」メディア（課題番号 24680078）の助成を受けて行った．

文　　献

[1] K.Aizawa, S.Kawasaki, T.Ishikawa, and T.Yamasaki, “Capture and retrieval of life log”, in Proc. of Int. Conf. on Artificial Reality and Telexistence (ICAT), pp. 49-55, 2004.

[2] B. H. Prananto, I. Kim, and H. Kim, “Multi-level Experience Retrieval for the Personal Lifelog Media System”, in Proc. of Third Int. IEEE Conf. on Signal-Image Technologies and Internet-Based System (SITIS), 2007.

[3] 近藤一晃, 高瀬恵三郎, 小泉敬寛, 中村裕一, 森幹彦, 喜多一, “個人視点映像を用いた気づき体験の回想と整理支援 -フィールド調査における問題発見を通じて-”, 電子情報通信学会研究会報告, PRMU2010-128, pp. 13-18, 2010.

[4] 岡田 昌也, 鳥山 朋二, 多田 昌裕, 角 康之, 間瀬 健二, 小暮 潔, 萩田 紀博, “実世界重要体験の抽出・再現に基づく事後学習支援手法の提案”, 電子情報通信学会論文誌 D-II, vol 91, no. 1, pp. 65-77, 2008.

[5] A. R. Doherty, A. F. Smeaton, K. Lee, and D. P. Ellis, “Multimodal Segmentation of Lifelog Data”, In Proc. on RIAO2007 - Large-Scale Semantic Access to Content, 2007.

[6] 井手一郎, 山本晃司, 浜田玲子, 田中英彦, “ショット分類に基づく映像への自動的索引付け手法”, 電子情報通信学会論文誌 D-II, vol. 82, no. 10, pp. 1543-1551, 1999.

[7] H. Luo, J. Fan, J. Yang, W. Ribarsky, and S. Satoh, “Exploring Large-Scale Video News via Interactive Visualization”, IEEE Symposium On Visual Analytics Science And Technology (VAST2006), pp. 75-82, 2006.

[8] 笠松沙紀, 伊藤貴之 “動画像データの要約可視化インタフェースの一手法”, DEIM2010 第 2 回データ工学と情報マネジメントに関するフォーラム, E-9, 2010.

[9] C. Barnes, D. B. Goldman, E. Shechtman, and A. Finkelstein, “Video tapestries with continuous temporal zoom”, Proc. of ACM SIGGRAPH2010, Vol. 29 Issue 4, No. 89 ,2010.

[10] 松井 研太, 近藤 一晃, 小泉 敬寛, 中村裕一, “輝度値の分布と情報量を用いた画像貼り合わせの評価”, 電子情報通信学会研究会報告, PRMU2013-32, pp. 77-82, 2013.

[11] 松井 研太, 近藤 一晃, 小泉 敬寛, 中村裕一, “個人視点映像からの広視野画像の自動生成 -輝度値の確率分布に基づいた貼り合わせに適した画像群の選択-”, 電子情報通信学会研究会報告, MVE2013-21, pp. 17-22, 2013.

[12] Y. Y. Schechner and S. K. Nayar, “Generalized Mosaicing: Polarization Panorama”, IEEE transactions on PAMI, Vol. 27, No. 4, pp. 631-636.

[13] A. Sibiryakov and M. Bober, “Graph-based multiple panorama extraction from unordered image sets”, in Proc. of SPIE 6498, Computational Imaging V, 2007.

[14] M.Brown and D.G.Lowe, “Recognising Panoramas”, In Proc. of the 9th International Conference on Computer Vision. Nice, vol. 2, pp. 1218-1225, 2003.

[15] M. Brown and D. Lowe., “Automatic Panoramic Image Stitching using Invariant Features”, International Journal of Computer Vision. vol. 74, No. 1, pages 59-73, 2007.

[16] Michael E.Tipping and Christopher M.Bishop, “Bayesian Image Super-resolution”, In NIPS 15, MIT Press 2003, pp. 1279-1286.

[17] 兼村 厚範, 前田 新一, 福田 航, 石井 信, “不確実性を手なずけるベイズ統計推測による画像超解像”, 電子情報通信学会誌, vol. 93, no. 9, pp. 759-763, 2010.

[18] G. K Chantas, N. P Galatsanos, N. A Woods, “Super-resolution based on fast registration and maximum a posteriori reconstruction”, IEEE Transactions on Image Processing, vol. 16, no. 7, pp. 1821-1830, 2007.

[19] 柳生健志, 川本一彦, “超解像処理のためのベイズ型情報量基準に基づく正則化パラメータの自動決定”, 第 24 回ファジィシステムシンポジウム講演論文集, 2D1-02, 2009.