H-059

複数モーション領域を含む時系列画像の分割レジストレーション

Image Registration Technique for Sequential Images with Multiple Motion Regions

張 馴權 †, 清水 雅夫 †, 奥富 正敏 †
SoonKeun Chang, Masao Shimizu, and Masatoshi Okutomi

1 まえがき

時系列画像中のある基準画像に対して入力画像の位置を合わせる画像レジストレーションは，画像の超解像処理や高画質化処理，画像モザイキング，トラッキングなどの多くの画像処理において，最も基本的かつ重要な処理である．多くの場合，レジストレーションでは，注目領域 (ROI: Region of Interest) 内の全ての画素が想定した単一のモーションモデルで表せることを前提としている．例えば，モーションモデルとして平面射影変換を想定すると，注目領域内の各画素は 3 次元空間中の同一平面に対応している必要がある．しかし，多くの自然画像では，このような理想的な状況は極めてまれで，注目領域内に複数のモーション領域が存在したり，想定したモーションで表せる領域が限定されていることが多い．このために，推定したモーションが不正確になることや，レジストレーション自体が失敗する可能性があった．

従来，画像中から同一モーション領域を抽出する手法として，学習を用いて背景画像を推定し，それとの差分により検出する方法 [5] や，密なオプティカルフローを用いたモーションセグメンテーション [2][3][4][6][7] 等が提案されてきた．しかし，学習による背景画像推定手法では固定カメラで撮影した画像を用いることを前提としているため，手持ちカメラを含む一般的な時系列画像に適用することが難しい．また，密なオプティカルフローを利用する手法では，対象の移動が大きくかつ剛体で近似できるときには有効だが，それ以外のときは対象の一部だけが注目領域として検出されるなどの問題点があった．ロバスト推定と k-means クラスタリングを使ったモーションセグメンテーションによるレイヤー表現 [6] は，時系列画像中の特定物体を消去する処理には有効だが，超解像処理にも利用可能なモーションとマスクを生成することはできない．

本論文では，手持ちカメラなどで撮影した複数のモーション領域を含む時系列画像から，各モーション領域を自動的に抽出し，基準画像に対して高精度にレジストレーションする手法を提案する．各モーション領域の抽出には，モーションパラメータを推定するときに有効な画素だけを選択して利用する手法 [8] を拡張して利用する．まず最初に，画像全体を注目領域に設定して，この中から同一モーション領域とその領域に対するモーションパラメータを同時に推定する．このときに抽出する領域は，同一のモーションで表せて，しかも面積が大きくリッチなテクスチャを持つ領域である [3]．次に，この抽出した領域を除いた領域に対して同様の処理を繰り返す．このように，順次同一モーション領域を抽出することで，時系列画像に含まれる全てのモーション領域を分割し，各領域に対して正確に推定したモーションパラメータを用いたレジストレーションが可能になる．

本論文は，以下のように構成する．第 2 節では，基準画像のある注目領域に対して，他の 1 枚の入力画像中から同一モーション領域とそのモーションパラメータを求める手法を述べる．第 3 節では，時系列画像の各入力画像中から，複数のモーション領域とそれぞれのモーションパラメータを順次求める手法を述べる．第 4 節では，提案手法を用いた 2 種類の実験結果を示す．

† 東京工業大学 大学院理工学研究科 機械制御システム専攻．
なお，第 1 著者の現在の所属は SAMSUNG TECHWIN CO., LTD.

2 同一モーション領域の抽出

2.1 モーションモデル

基準画像 $I_0(\mathbf{x})$ 上に設定した注目領域 (ROI) に対して，入力画像 $I_t(\mathbf{x})$ をレジストレーションするときの目的関数を，次のように定義する．

$$E(\mathbf{h}_t) = \sum_{\mathbf{x}\in \mathrm{ROI}} \mathbf{M}_t(\mathbf{x}) |I_t(\mathbf{W}(\mathbf{x};\mathbf{h}_t)) - I_0(\mathbf{x})|^2 \quad (1)$$

ただし，$I_t(\mathbf{W}(\mathbf{x};\mathbf{h}_t))$ は基準画像に合わせるようにアフィン変換した入力画像を表す．また，$\mathbf{x} = [x, y, 1]^\top$ と $\mathbf{W}(\mathbf{x};\mathbf{h}) = \begin{bmatrix} h_1 & h_2 & h_3 \\ h_4 & h_5 & h_6 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}$ は，変換前と変換後の画像座標の同次座標系表現，$\mathbf{h} = [h_1, h_2, h_3, h_4, h_5, h_6]^\top$ はアフィンモーションを表す 6 個のパラメータベクトルである．$\mathbf{M}_t(\mathbf{x})$ は，レジストレーションに適切な領域を表すマスク画像である．式 (1) は，Gauss-Newton 法などを利用して最小化することができる [1]．

2.2 マスク画像とモーションパラメータの同時推定

基準画像上に設定した注目領域に対して，入力画像のモーションパラメータがほぼ正確に求まったと仮定する．

このとき，各画像の注目領域内からそれぞれ小領域 (Patch) を取り出し，次のように ZNSSD(Zero-mean Normalized Sum of Squared Differences) 値を計算する．

$$R(\mathbf{x},\mathbf{u}) = \frac{1}{\sigma_{I'}\sigma_{I_0}} \sum_{\mathbf{x}\in \mathrm{Patch}} |\{I'(\mathbf{x}+\mathbf{u}) - \mu_{I'}\} - \{I_0(\mathbf{x}) - \mu_{I_0}\}|^2 \quad (2)$$

ここで，$I'(\mathbf{x}) = I_t(\mathbf{W}(\mathbf{x};\mathbf{h}))$ は入力画像を推定したモーションパラメータで変換した画像を表す．$\mathbf{u} = [u, v, 1]^\top$ は 2 次元の平行移動を表すベクトルである．$\mu_{I'}$ と μ_{I_0}，及び $\sigma_{I'}$ と σ_{I_0} は，それぞれ画像 $I'(\mathbf{x})$ と基準画像 $I_0(\mathbf{x})$ 上の小領域内の輝度値の平均と標準偏差を表す．

$(u, v) = \{(-1, 0), (0, 0), (1, 0), (0, -1), (0, 1)\}$ の 5 個所の位置に対して式 (2) の値を計算し，値の分布を調べることで，注目領域内の小領域が正確にレジストレーションできているこ

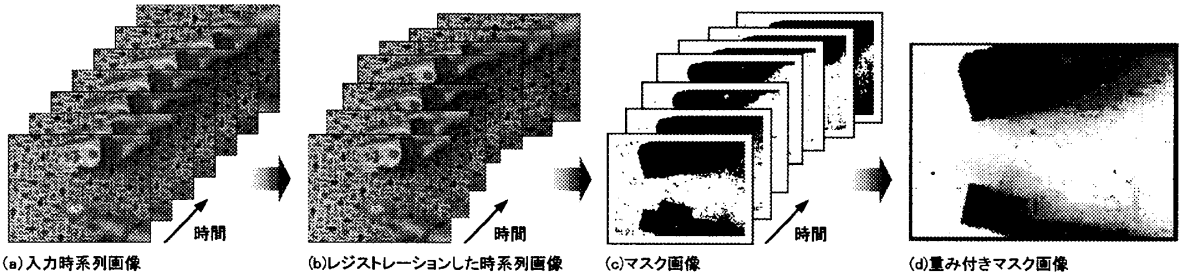

図 1　背景領域に対するマスク画像と重み付きマスク画像.

とを検出することができる [8]. 小領域が正確にレジストレーションされていたときに，マスク画像を $M_t(\mathbf{x}) = 1$ として，それ以外は 0 とする．このようにして得られたマスク画像は，レジストレーションされた注目領域内で正確に位置が合っている画素を表している．

さらに，得られたマスク画像を使って，式 (1) の目的関数を最小化するモーションを再推定する．このように，マスク画像とモーションが収束するまで繰り返す [8]. モーションの推定にも同様に ZNSSD を利用するために，得られたマスク画像を使って注目領域内の平均と標準偏差を求め，基準画像と入力画像をそれぞれ正規化する．

物体の明るさが変化するときにも安定にマスク画像を得るためには，式 (2) の代わりに正規化相互相関も利用できる．しかし，勾配法を利用したモーション推定でも同じ正規化を行う必要があるため，勾配法でも利用可能な ZNSSD を利用した．

3　複数モーション領域の抽出

3.1　重み付きマスク画像

時系列画像中の各入力画像に対して，同一モーション領域を表すマスク画像 $\mathbf{M}_t(\mathbf{x})$ が全て得られたとする．このマスク画像は，ある 1 枚の入力画像に対する同一モーション領域を表しているが，ノイズやテクスチャの影響で対象の一部が欠けたり，対象以外の領域が含まれたりすることがある [3].

そこで，このマスク画像の時間方向の総和をとり，各位置に対する重みを 0 から 1 に正規化することで，基準画像に対して同一モーション領域を表す重み付きマスク画像 $\mathbf{C}(\mathbf{x})$ を生成する．ただし，マスク画像 $\mathbf{M}_t(\mathbf{x})$ の面積がしきい値 $T_{\mathrm{fail}}(\approx 400)$ 以下のときには，そのマスク画像は加算しない．マスク画像の面積が小さくなると，レジストレーションは不安定になる．このしきい値は，安定なレジストレーション結果をもとに得られたマスク画像だけを利用するために，最終的に安定なレジストレーションが可能な最小値として実験的に決定した．

図 1 に，複数モーション領域を含む時系列画像と，背景領域に対するマスク画像及び重み付きマスク画像を示す．この時系列画像（図 3 と同じ）は，両手に持った茶箱とカードを動かしている様子を手持ちカメラで撮影したものである．同図 (c) に基準画像（先頭フレーム）に対して抽出したマスク画像（白い部分が抽出された同一モーション領域）を，同図 (d) に重み付きマスク画像を示す．

3.2　複数モーション領域抽出アルゴリズム

3.1 では，画像中で面積が大きくリッチなテクスチャを持つ領域を，マスク画像 $\mathbf{M}_t(\mathbf{x})$ 及び重み付きマスク画像 $\mathbf{C}(\mathbf{x})$ として抽出する方法を説明した．以下で，この方法を順次使って，画像に含まれる複数モーション領域を全て抽出するアルゴリズムを示す．

1. モーション領域番号を $i \leftarrow 1$ に初期化する．
2. モーション領域番号 i に対する，次の重み付き有効領域画像 $\mathbf{A}_i(\mathbf{x})$ を求める．

$$\begin{aligned} \mathbf{A}_1(\mathbf{x}) &= 1 \\ \mathbf{A}_i(\mathbf{x}) &= \mathbf{A}_{i-1}(\mathbf{x})(1 - \mathbf{C}_{i-1}(\mathbf{x})) \quad \text{for } i \geq 2 \end{aligned} \tag{3}$$

画像中の $\mathbf{A}_i(\mathbf{x})$ で表す有効領域の中から，モーション領域とモーションパラメータを抽出する．ただし，$\mathbf{A}_i(\mathbf{x})$ の面積が一定値以下のときには，全てのモーション領域の抽出が終了したと判断して計算を終了する．

3. 有効領域画像 $\mathbf{A}_i(\mathbf{x})$ を利用して，次式を最小化するモーションパラメータとマスク画像を各フレームに対して推定する．

$$E(\Delta\mathbf{h}_t) = \sum_{\mathbf{x}\in\Omega} \mathbf{A}_i(\mathbf{x})\mathbf{M}_{t,i}(\mathbf{x})|I_t(\mathbf{W}(\mathbf{x};\mathbf{h}_{t-1,i} + \Delta\mathbf{h}_{t,i})) - I_0(\mathbf{x})|^2 \tag{4}$$

ただし，Ω は画像全体を表す．また，勾配法の初期値には前フレームで推定したモーションパラメータを使用する．

4. モーション領域番号 i に対するマスク画像 $\mathbf{M}_{t,i}(\mathbf{x})$ から，重み付きマスク画像 $\mathbf{C}_i(\mathbf{x})$ を求める．
5. モーション領域番号を $i \leftarrow i + 1$ に更新して，2. に戻る．

図 2 に，重み付き有効領域画像 $\mathbf{A}_i(\mathbf{x})$ と，重み付きマスク画像 $\mathbf{C}_i(\mathbf{x})$ の例を示す．左の列 ($i = 1$) では，画像全体の中から背景領域が重み付きマスク画像として抽出されている．中の列 ($i = 2$) では，背景領域を除いた領域から下側のモーション領域（カード）が抽出されている．さらに，右の列 ($i = 3$) では，背景領域とカード領域を除いた領域から，上側のモーション領域（茶箱）が抽出されている．

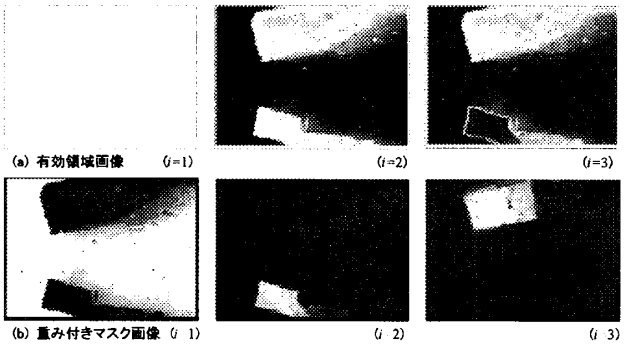

図 2　複数モーション領域の抽出.

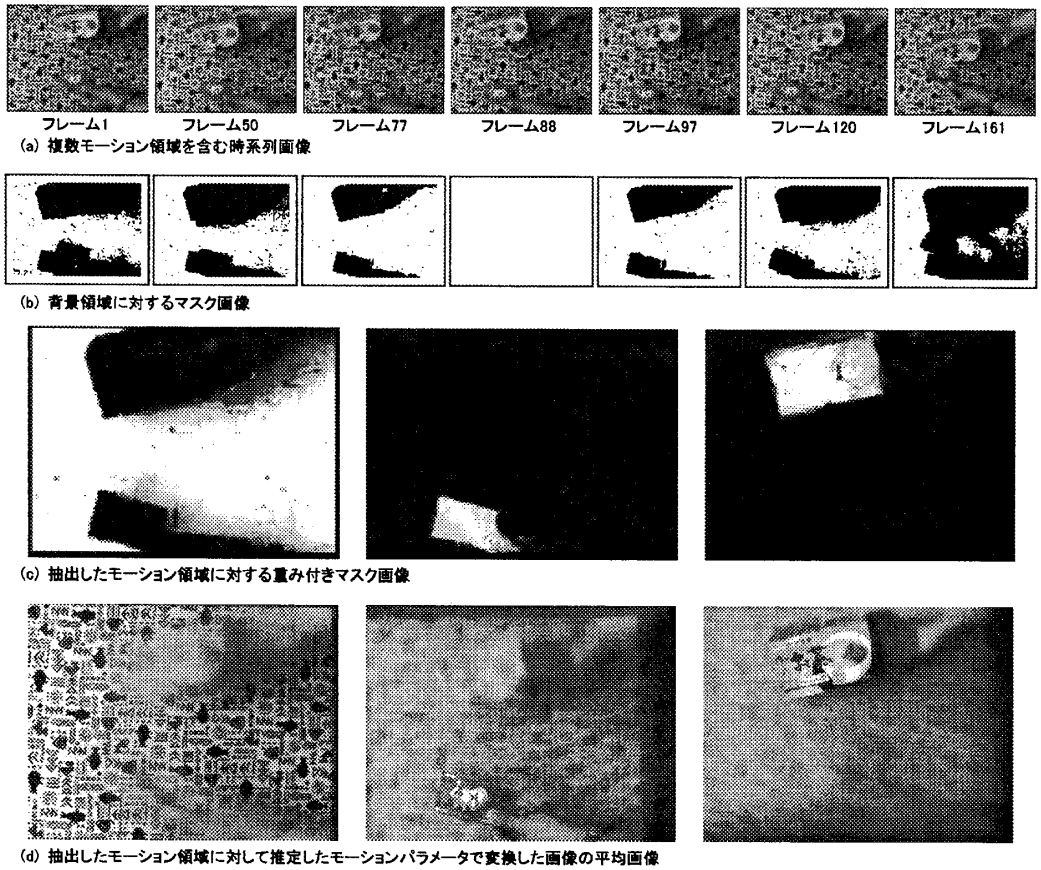

図 3　複数モーション領域の抽出と変換画像を使った平均画像.

4 実験結果

4.1 複数モーション領域の抽出とレジストレーション

図 3(a) に，実験に使用した時系列画像の一部を示す．この時系列画像は 200 フレームで構成され，Point Grey 社製の Dragonfly カメラ（単板カラー CCD，画像サイズ 640×480[画素]，30FPS）で，両手に持った茶箱とカードを動かしている様子を手持ちカメラで撮影したものである．線形補間法でデモザイキング処理を行った後，輝度成分を利用してレジストレーションを行った．

図 3(b) に，フレーム 88 を基準画像としたときの背景領域に対するマスク画像を示す．背景領域に対して異なるモーションを持つ領域は，黒くマスクされていることがわかる．この背景領域に対するマスク画像から求めた重み付きマスク画像が，同図 (c) 左である．

重み付きマスク画像から順次有効領域画像を求めて，画像に含まれるその他のモーション領域を抽出した結果が，図 3(c) である．この時系列画像では，最終的に 3 つのモーション領域が検出された．

基準画像に対して抽出したモーション領域に対して，各フレームのモーションパラメータが推定済みなので，このモーションパラメータを使って各フレームを変換すれば基準画像にレジストレーションすることができる．図 3(d) に，このようにして各フレームを変換してから求めた平均画像を示す．抽出したモーション領域では各フレームが正確にレジストレーションされているために画像がぼやけず，それ以外の領域はぼやけた結果になる．

4.2 時系列画像中の特定物体消去

4.1 の実験では，抽出したモーション領域に対して推定したモーションパラメータを使って，入力画像全体を変換してから平均画像を作成した．このとき，入力画像全体ではなく，抽出したモーション領域だけを使って平均画像を作成すれば，他のモーション領域を消去した平均画像を作成できる．さらに，基準画像を順次変更しながらこの処理を行えば，最終的には全てのフレームを基準画像としたときの結果が得られる．以上の処理を利用すれば，もともとの時系列画像と同じ視点でありながら，特定のモーション領域だけを消去することができる．

図 4(a) に，実験に使用した時系列画像 (flower garden) の一部を示す．20 フレームを使用して実験を行った．画像サイズは 352 × 240[画素] である．

図 4(b) に，第 11 フレームを基準画像としたときに抽出された 3 つのモーション領域（花壇，家，及び木）を示す．

図 4(c) に，各入力フレームを基準画像としたときに抽出した「花壇」領域を示す．基準画像を先頭フレームから最終フレームまで順次変更しながら，提案手法によって複数モーション領域抽出とモーションパラメータ推定を行った．さらに，重み付きマスク画像でマスクすることで，「花壇」領域の画像を抽出した．

図 4(d) に，各基準画像に対して推定したモーションパラメータを使って変換した，他のフレーム画像中の「花壇」領域を全て平均した画像を示す．「花壇」領域だけを平均するので，画像中から抽出されていない「家」領域や「木」領域が消去されている．しかも，各フレーム画像を順次基準画像として設定するので，もとの画像と同じ視点の時系列画像が得られている．

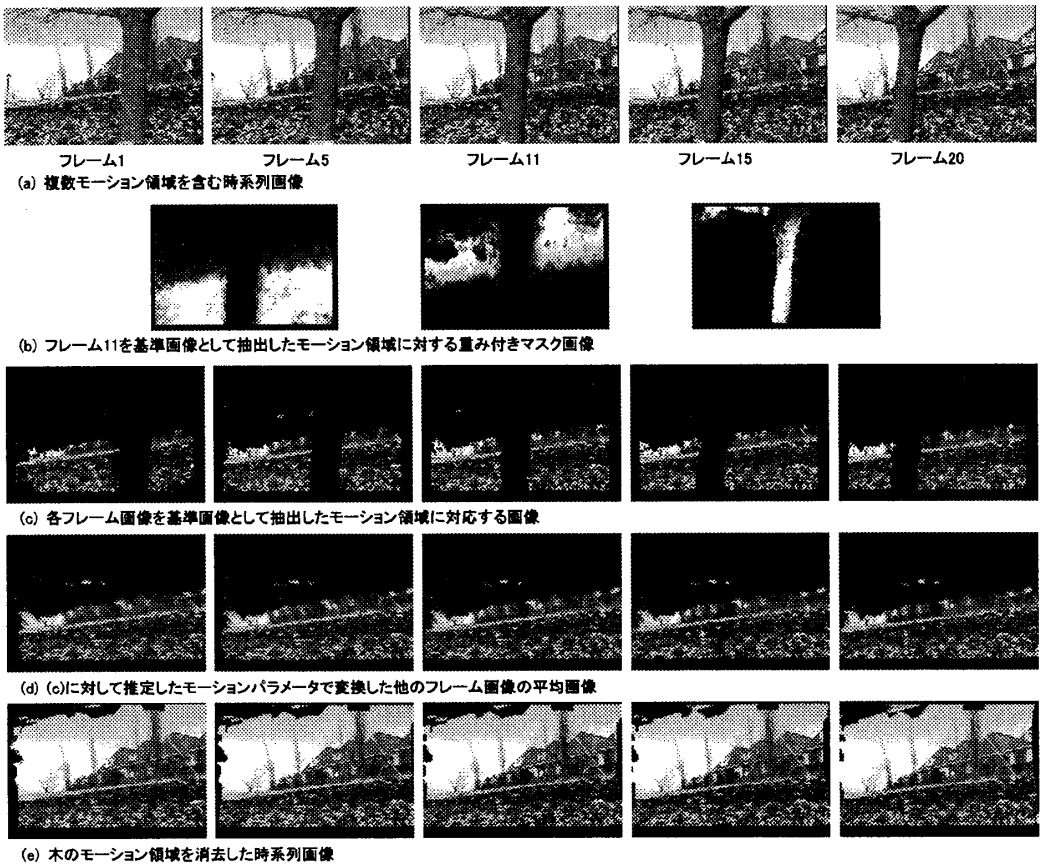

図 4　複数モーション領域抽出結果を利用した特定物体の消去.

図 4(e) に，このような処理を「家」領域に対しても行い，「花壇」と合わせて平均した画像を示す．入力時系列画像（同図 (a)）から，「木」領域だけを消去した時系列画像が生成できている．同図 (e) 中の黒で示す領域は，「木」領域によって常に遮蔽されていた領域である．

5 まとめ

手持ちカメラなどで撮影した複数のモーション領域を含む時系列画像から，各モーション領域を自動的に抽出して基準画像に対して高精度にレジストレーションする手法を提案した．提案手法では，モーションパラメータ推定をするときに有効な画素だけを選択して利用する手法を拡張して利用した．提案手法を用いると，時系列画像に含まれる全てのモーション領域を分割して，しかも各領域に対して正確に推定したモーションパラメータを用いたレジストレーションが可能になる．

提案手法を利用して抽出したモーション領域とモーションパラメータを使えば，入力画像全体ではなく，抽出したモーション領域だけを使って平均画像を作成することができる．この結果，他のモーション領域を消去した平均画像を作成できる．さらに，基準画像を順次変更しながらこの処理を行えば，最終的には全てのフレームを基準画像としたときの結果が得られる．すなわち，もともとの時系列画像と同じ視点でありながら，特定のモーション領域だけを消去することができる．

実画像を使った実験を行い，提案手法の有効性を確認した．

参考文献

[1] S. Baker and I. Matthews, "Lucas-Kanade 20 Years On: A Unifying Framework", International Journal of Computer Vision, vol. 56, no. 3, pp. 221–255, 2004.

[2] M. J. Black and P. Anandan, "The Robust Estimation of Multiple Motions: Parametric and Piecewise-Smooth Flow Fields", Computer Vision and Image Understanding, vol. 63, no. 1, pp. 75–104, 1996.

[3] M. Irani, B. Rousso, and S. Peleg, "Computing Occluding and Transparent Motions", International Journal of Computer Vision, vol. 12, no. 1, pp. 5–16, 1994.

[4] G. Piriou, P. Bouthemy, and J.-F. Yao, "Learned Probabilistic Image Motion Models for Event Detection in Videos", IEEE Proc. on International Conference on Pattern Recognition, vol. 4, pp. 207–210, 2004.

[5] Y. Sheikh and M. Shah, "Bayesian Modeling of Dynamic Scenes for Object Detection", IEEE Trans. on Pattern Analysis and Machine Intelligence, vol. 27, no.11, pp. 1778–1792, 2005.

[6] J. Y. A. Wang and E. H. Adelson, "Representing Moving Images with Layers," IEEE Trans. on Image Processing, vol. 3, no. 5, pp. 625–638, 1994.

[7] L. Wixson, "Detecting Salient Motion by Accumulating Directionally-Consistent Flow", IEEE Trans. on Pattern Analysis and Machine Intelligence, vol. 22 no. 8, pp. 774–780, 2000.

[8] 張 馴樘, 清水 雅夫, 奥富 正敏, "領域選択を伴う 2 段階レジストレーション", 電子情報通信学会論文誌 D, vol. J90-D, no. 2, pp. 514–525, 2007.