

動画編集のための動き推定

鈴木 裕^{†1} 山口 泰^{†1}

動画編集の自動化には、時刻が異なるフレーム間における物体の対応関係が必要となる。ところが、大きな動きやオクルージョンなどの影響で、一般的な動き追跡の方法では必要な対応関係が得られないことがある。本研究では、大きな動きでも破綻することなく、正しい動きとオクルージョン領域を推定できる方法を考える。

Motion estimation for video editing

YUICHI SUZUKI^{†1} and YASUSHI YAMAGUCHI^{†1}

For automation of video editing, the correspondence relation of the objects at different time points is required. However common object tracking methods fail to acquire correct relations due to large motions and occlusion. In this research, we suggest the method for estimating correct motions and occluded areas in the case that there is any large motion.

1. はじめに

動画は基本的に多数の連続した静止画像の集まりである。よって、動画の編集作業は、静止画像の編集作業と比較して作業量が莫大になりがちである。例えば、特定物体を動画の中から抽出・削除する、特定物体に模様を描く、などといった作業は基本的に多くのフレームに対して手作業で処理を施さなくてはならない。こういった処理を自動的に行うには、時間的に異なるフレームに存在する同一物体間の対応関係をとる必要がある。

予め計算された対応関係を利用した研究としては Alex らの研究¹⁾ が挙げられる。Alex らの研究では、パーティクルビデオ²⁾ を利用して、動画中の物体の動きを粒子の座標で表

現している。物体の運動が粒子で表現されているということは、時刻 t_α で位置 p^α 、時刻 t_β で位置 p^β の粒子 $Particle_A$ の位置が、

$$P_A(t_\alpha) = p^\alpha, P_A(t_\beta) = p^\beta$$

と表現できるような関数 $P_A(t)$ が存在するということである（オクルージョンは起きていないとする）。つまり、特定の粒子が異なる時刻にどの位置に存在するかが容易に分かるといことである。このような性質を利用して Alex らは動画中の 3 次元物体の表面を 2 次元画像（モザイク画像）に展開し、その 2 次元画像に編集を加えることで、動画中の物体に編集が反映されるシステムを開発した。

パーティクルビデオでは粒子の運動を計算するのに、標準的なオプティカルフローの計算を利用している。ここでいう標準的なオプティカルフローとは、

動画 I

時空間立体における位置 $\mathbf{x} = (x, y, t)$

オプティカルフロー $\mathbf{u}(\mathbf{x}) = (u(\mathbf{x}), v(\mathbf{x}), \Delta t)$

を用いて表されるエネルギー汎関数 E を最小化することで求められるものである。 E は例えば、

$$\begin{aligned} E(\mathbf{u}) &= E_D(\mathbf{u}) + \lambda E_S(\mathbf{u}), \\ E_D(\mathbf{u}) &= \int_{\Omega} \Psi \left((I(\mathbf{x} + \mathbf{u}) - I(\mathbf{x}))^2 \right) d\Omega, \\ E_S(\mathbf{u}) &= \int_{\Omega} \Psi \left(\|\nabla \mathbf{u}\|^2 \right) d\Omega, \\ \Psi(x^2) &= \sqrt{x^2 + \epsilon} \end{aligned}$$

といった形で定義される。ここで E_D はオプティカルフローが画像データに適合していると小さくなるように定義されたエネルギー（データエネルギー）で、 E_S はオプティカルフローが滑らかに分布することを強制するエネルギー（スムーズエネルギー）である。 λ が大きくなればなるほど、滑らかなオプティカルフローの分布が得られる。

ここで定義されたエネルギー E を最小化するような \mathbf{u} を求める場合、変分法によって得られるオイラーラグランジュ方程式を解くことが多い（最急降下法）。ところが、

$$I(\mathbf{x} + \mathbf{u}) - I(\mathbf{x})$$

という部分は非常に非線形であるため、実際に数値計算する場合は 1 次近似により

$$\frac{\partial I}{\partial x} u + \frac{\partial I}{\partial y} v + \frac{\partial I}{\partial t} \Delta t$$

^{†1} 東京大学
The University of Tokyo

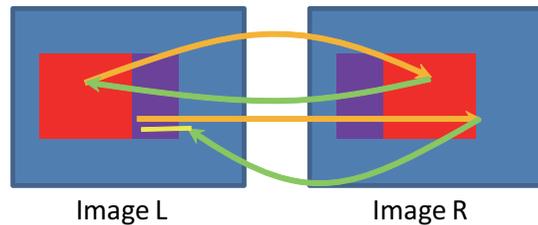


図 1 ディスパリティエラー．黄色の部分がエラーとなる．
Fig.1 The yellow line indicates disparity error.

と表されることが多い．これは， u が非常に小さい場合にのみ良い近似を与える．また，最急降下法は基本的に局所解に収束するので，しばしばピラミッドイメージを使って計算される．

上記のような方法では，対応する点が存在しない領域（オクルージョン領域）は考慮されないか，考慮されていても明示的にエネルギーの式に関係しない．なぜなら，真のオプティカルフローが十分に小さければ，オクルージョン領域も十分に小さく，計算結果に無視できる程度の誤差しか与えないからである．多くのオプティカルフローの研究では，そういった十分小さいフローを持つテストセットが使用されているので，あまりオクルージョンを考慮されないが，一般の動画像では示すような，大きな動きが現れることがある．よって，動画像の編集のためにオプティカルフローを用いて対応関係を計算したい場合，大きな動きが現れた場合でも計算が破綻しないような方法を用いる必要がある．パーティクルビデオでは，標準的な方法で計算されたオプティカルフローを信頼して粒子の動きを計算しているので，大きな動きがあるとうまく計算できない．

2. 関連研究

オクルージョンを考慮した最近の研究としては，変分法を用いた物³⁾⁴⁾，や離散 MRF (Markov Random Field) を用いた物⁵⁾がある．例えば Occlusion Aware Optical Flow⁴⁾では，図 1 に示すような，ディスパリティエラーを用いている．左の絵からの対応関係が，オレンジ色の矢印で示され，右の絵からの対応関係が黄緑色で示されている．このとき，紫色のオクルージョン領域から出ている，オレンジ色の矢印の先からもどってくる黄緑色の矢印はもとの矢印の根本と一致しない．黄色の線分だけずれが生じたことになり，このずれが大きいほど，オクルージョン領域である可能性が大きくなる．このディスパリティエラーを

利用することで，オクルージョンエネルギーを定義し，オクルージョンエネルギーを含む全体のエネルギーを最小化することでオクルージョン領域が与える誤差が，対応点が存在する領域に影響しないように工夫している．離散 MRF を使用した場合，オプティカルフローの計算が，有限のラベルを割り当てる問題に変換されるので，取り得る値が離散的な値になるという欠点があるが，エネルギー汎関数の定義が柔軟に設定できるという利点がある．

オプティカルフローとは少々異なるが，オクルージョンを非常に厳密に扱うという意味ではステレオマッチングの分野の研究も関連する⁶⁾⁷⁾，ステレオの場合，物体の変形や，垂直方向の変位は考えないので，一般的なオプティカルフローよりも制約は強くすることができる．その分，視差の大きなところなどでも比較的正確な対応関係とオクルージョン判定が実現できている．

本研究では，Occlusion Aware Optical Flow および一般的な高精度オプティカルフローとして Brox らの方法⁸⁾，そして対称的ステレオマッチング⁷⁾をオプティカルフローに適用できるようにした方法を比較し，大きな動きがうまく計算できるかを検証する．

本研究報は以下の構成である．第 3 節では対称的ステレオマッチングの方法について説明し，第 4 節で合成画像および自然画像に対する実験を行う．そして第 5 節で結論を，第 6 節において今後の展望を述べる．

3. 対称的ステレオマッチング

対称的ステレオマッチングでは，2 枚の画像 I_L, I_R について，最小化すべきエネルギーが対称的に構築されている．つまり，画像 I_L からみたエネルギー汎関数 E_L と，同様に構築された I_R からみたエネルギー汎関数 E_R の両方を最小化する．ここでは E_L について簡単に説明するが， E_R についてもほぼ同様に計算される．以下エネルギー汎関数のデータ項，スムーズ項，可視性項について説明する．

3.1 データ項

まず記号を定義する．

$O: \{0, 1\} \in O$ オクルージョンラベル．オクルージョン領域は 1．そうでない場合は 0．

$D: d \in D$ ディスパリティラベル．本研究ではオプティカルフローと同義．

$F(x, d)$ データマッチングコスト．対応点のデータが一致していればいほど小さな値を返すような関数．

データ項とは、対応点どうしの輝度値（色）の一貫性を表現する項である。データ項は

$$E_d(D_L, O_L) = \sum_{O_L(\mathbf{x})=0} F(\mathbf{x}, \mathbf{d}) + \sum_{O_L(\mathbf{x})=1} \eta_0$$

と表される。 η_0 はオクルージョンコストで、全体がオクルージョンと判定されるのを防ぐ。つまり、オクルージョンが発生している部分では対応関係がとれないので、その部分では一定のコストがかかると定義してある。

3.2 スムーズ項

スムーズ項では、隣り合ったピクセルのラベルが、類似したラベルになるように定義される。つまり、ピクセル i の隣接ピクセル集合 N に関して、

$$E_s(D_L) = \sum_i \sum_{j \in N} distance(d_i, d_j),$$

と定義できる。 $distance$ はラベルの距離を定義する関数。

3.3 可視性項

可視性項は、可視性に関する一貫性を表現する。つまり、画像 I_L のオクルージョン領域は、画像 I_R から見た対応関係が存在してはいけないということを強制する。 $W_L(D_R)$ を I_R から見た対応点が存在する場合 0、そうでない場合 1 とすると、

$$E_v(O_L, D_R) = \beta_w \sum |O_L(\mathbf{x}) - W_L(\mathbf{x}; D_R)| + \beta_o \sum_i \sum_{j \in N} |O_L(i) - O_L(j)|$$

となる E_v が定義できる

3.4 最適化

左側の画像だけで、オクルージョンラベルと、ディスパリティラベルという 2 種類の MRF を利用することになる。これらは相互に依存しているので、それぞれの MRF を解くときには、相手を固定して解く。これを何度か反復し、収束するまで計算する。

4. 実験

本研究では対称的ステレオマッチングのエネルギーを FastPD という方法⁹⁾¹⁰⁾ で最小化する。FastPD は主問題と双対問題を同時に解くことで高速で良い近似精度を持つ方法である。以降本研究の提案手法を SSM と表記する。一般的な変分法を用いるオプティカルフロー⁸⁾ を OF、オクルージョンを考慮したオプティカルフロー⁴⁾ を OAOF と表記する。SSM のオクルージョン領域は、オクルージョンラベルの値を見ている。OF のオクルージョン領域は、最終的に得られたフローのディスパリティエラーを使って計算している。OAOF は、エネルギーの式の中に陽に含まれているディスパリティエラーの最終的な値を表示している。

まず、3 種類の合成画像について評価する。それぞれ運動の大きさが小、中、大となる画像である。図 4 では (1) の画像を基準として (2,3,4) がそれぞれ小、中、大の運動に対応する。(5) はオプティカルフローの色表示であるカラーホイールである。(6) 以降は実験結果であるが (方法):(運動の大きさ):(画像の種類) という意味である。画像の種類は、L が基準で、R がもう一枚の画像である。Occ は L のオクルージョンマップで、黒い部分がオクルージョン領域である。

4.1 合成画像についての結果

まず動きが小の画像の結果を見る。SSM はほとんど真の動きと同じである。オクルージョンと判定された領域もほとんど正しい。OF は、フローの精度とオクルージョン領域の判定、どちらもかなり正しい。OAOF では、ややぼやけてしまっている。

動きが中の画像の結果を見る。SSM では、動き、オクルージョン判定の両方ともほとんど真の値と同じである。OF では、オクルージョンが発生している部分で非常にエラーが大きくなっている。OAOF の結果は一見正確に見えるが、この手法の特徴としてオクルージョン領域のエラーを伝播させないというものがあるため、動き境界でのエラーが少ないだけである。つまり、計算されたオプティカルフロー自体はあまり正確ではない。それはオクルージョン領域の判定に失敗していることから分かる。

動きが大きい画像の結果を見る。SSM では動き境界でわずかなエラーがみられるが、オクルージョン領域の判定と共に、かなり高精度に計算できている。OF では完全に計算が破綻している。OAOF では動きが中の画像と同様に、動き境界でのエラーはほとんど無いが、実際に計算されたオプティカルフローの値とオクルージョン領域はほとんど正しくない。

4.2 自然画像についての結果

自然画像として図 3 に示す、傘を振り回している動画像を使用した。各画像の情報は (方法):(画像の種類) で、L が L から見たオプティカルフロー、R が R から見たオプティカルフロー、OccL と OccR はそれぞれのオクルージョンラベルである。

SSM では、傘の部分の動きが正確に計算されていることが分かる。またオクルージョン領域も正しく判定できている。ただし、SSM の解は離散的な物なので、オプティカルフローがあまりスムーズではない。OF では、傘の部分がぼやけてしまって、動き境界でフローが不正確になっている。OAOF では OF よりも動き境界は正確だが、全体的に精度が足りなく、ぼやけてしまっている。

5. 結果の考察

SSM は全体的に良好な性能を示した。現在の実装では、フローのラベルが多くなると、メインメモリが足りなくなってしまうため、ラベルの種類を 150 種類程度までに制限している。またオプティカルフローのラベルをはじめから決めてしまっているため、オプティカルフローの精度に限界がある。しかし、SSM の結果は安定しており、動き境界が正確にとれている。一方 OAOF は、精度が不十分だが動き境界付近では比較的安定し、結果も滑らかになっている。これらの手法を組み合わせることで、より精度の高い結果を得ることができるとも思われる。

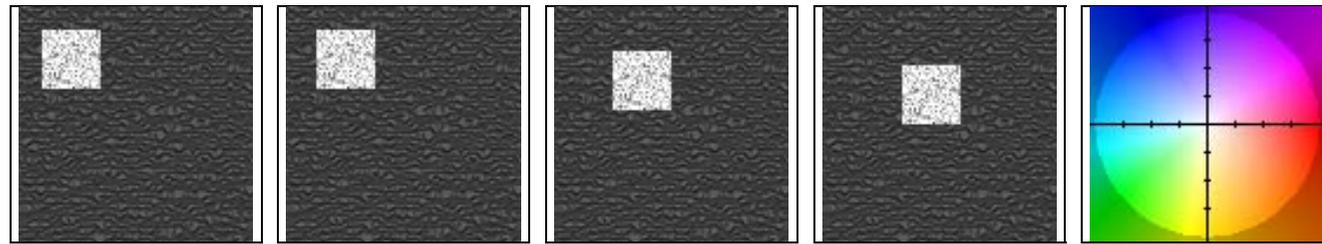
6. おわりに

本研究では、離散的な解法が大きな動きを持つ物体の動き境界を正確に計算するために利用できる可能性があることを示した。現段階では実装の問題により、大規模な MRF を解けないが、これを解決する方法は分かっており、今後さらにラベルの数を増やして、動的にラベルを決定するプログラムを使って実験を行う。また、連続的な解法と組み合わせることで、両方の長所を取り入れた解法を研究する予定である。

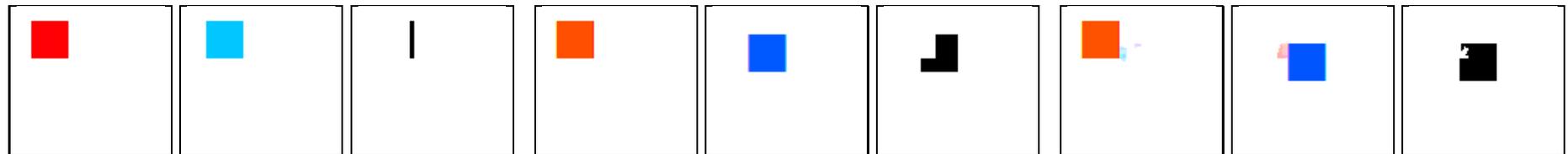
参 考 文 献

- 1) Rav-Acha, A., Kohli, P., Rother, C. and Fitzgibbon, A.: Unwrap mosaics: a new representation for video editing, *ACM Trans. Graph.*, Vol.27, No.3, pp.1–11 (2008).
- 2) Sand, P. and Teller, S.: Particle Video: Long-Range Motion Estimation using Point Trajectories, *Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on*, Vol.2, pp.2195–2202 (2006).
- 3) Alvarez, L., Deriche, R., Papadopoulos, T. and Sánchez, J.: Symmetrical Dense Optical Flow Estimation with Occlusions Detection, *International Journal of Computer Vision*, Vol.75, No.3, pp.371–385 (2007).
- 4) Ince, S. and Konrad, J.: Occlusion-Aware Optical Flow Estimation, *Image Processing, IEEE Transactions on*, Vol.17, No.8, pp.1443–1451 (2008).
- 5) Lim, K.P., Das, A. and Chong, M.N.: Estimation of occlusion and dense motion fields in a bidirectional Bayesian framework, *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, Vol.24, No.5, pp.712–718 (2002).
- 6) Zhang, L. and Seitz, S.M.: Estimating Optimal Parameters for MRF Stereo from a Single Image Pair, *IEEE Trans. Pattern Anal. Mach. Intell.*, Vol.29, No.2, pp.331–342 (2007).

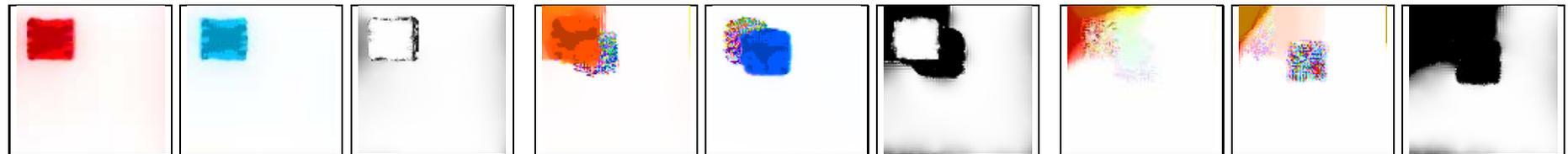
- 7) Sun, J., Li, Y., Kang, S. and Shum, H.: Symmetric stereo matching for occlusion handling, *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, Vol.2, pp.399–406 vol. 2 (2005).
- 8) Brox, T., Bruhn, A., Papenberg, N. and Weickert, J.: High Accuracy Optical Flow Estimation Based on a Theory for Warping, *Computer Vision - ECCV 2004*, pp.25–36 (2004).
- 9) Komodakis, N. and Tziritas, G.: Approximate Labeling via Graph Cuts Based on Linear Programming, *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, Vol.29, No.8, pp.1436–1453 (2007).
- 10) Komodakis, N., Tziritas, G. and Paragios, N.: Performance vs computational efficiency for optimizing single and dynamic MRFs: Setting the state of the art with primal-dual strategies, *Comput. Vis. Image Underst.*, Vol.112, No.1, pp.14–29 (2008).



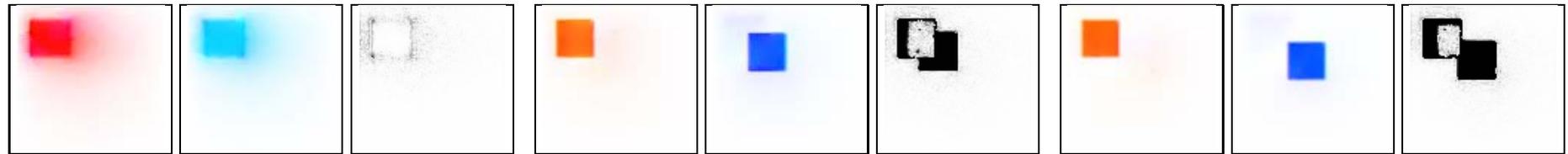
(1) 基準画像 (2) 運動：小 (3) 運動：中 (4) 運動：大 (5) カラーホイール



(6)SSM : 小 : L (7)SSM : 小 : R (8)SSM : 小 : Occ (9)SSM : 中 : L (10)SSM : 中 : R (11)SSM : 中 : Occ (12)SSM:大 : L (13)SSM : 大 : R (14)SSM : 大 : Occ



(15)OF : 小 : L (16)OF : 小 : R (17)OF : 小 : Occ (18)OF : 中 : L (19)OF : 中 : R (20)OF : 中 : Occ (21)OF:大 : L (22)OF : 大 : Occ (23)OF : 大 : R



(24)OAOF : 小 : L (25)OAOF : 小 : R (26)OAOF : 小 : Occ (27)OAOF : 中 : L (28)OAOF : 中 : R (29)OAOF : 中 : Occ (30)OAOF:大 : L (31)OAOF : 大 : R (32)OAOF : 大 : Occ

図 2 合成画像の実験結果 .
Fig.2 The result of composit images.

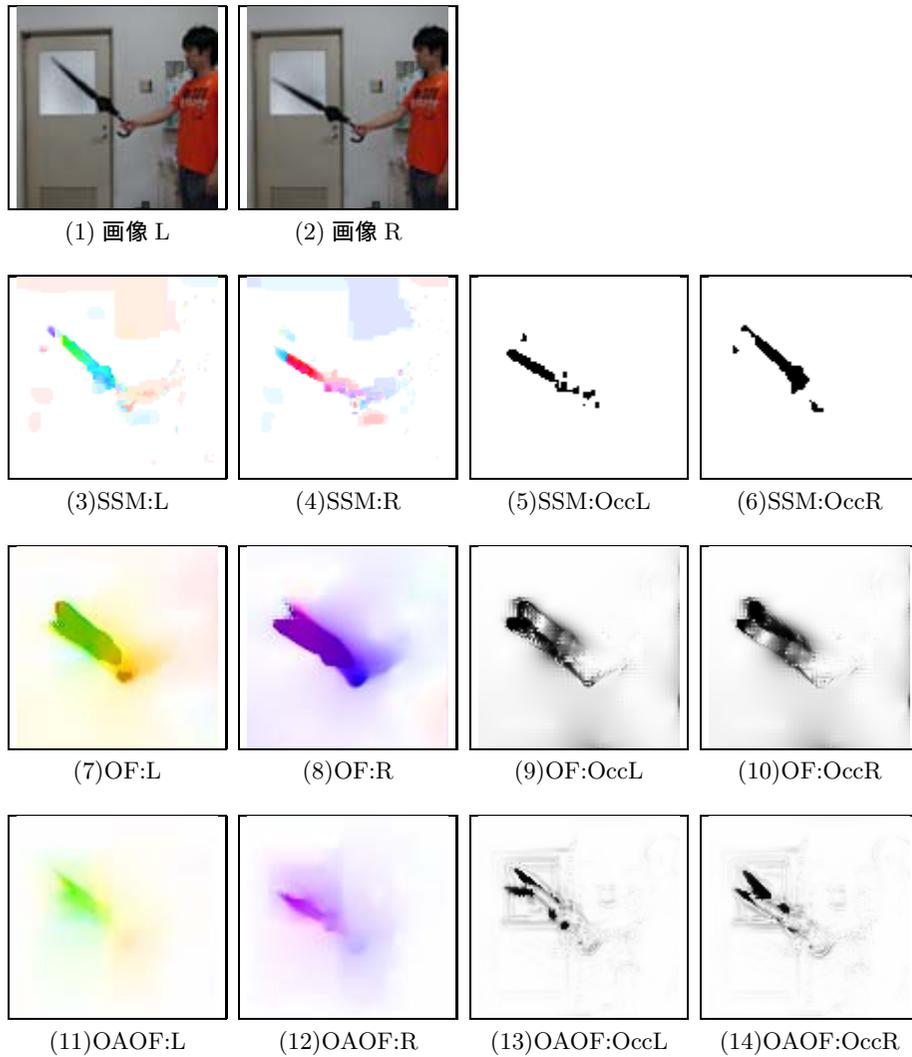


図 3 自然画像の結果 .
Fig. 3 The result of natural images.