

SIFT 特徴量に基づく Mean-Shift 探索による特徴点追跡

都築 勇司† 藤吉 弘亘† 金出 武雄††

† 中部大学 工学部 情報工学科

†† カーネギーメロン大学ロボティクス工学研究所

E-mail:tsuduki@vision.cs.chubu.ac.jp, hf@cs.chubu.ac.jp, tk@cs.cmu.edu

あらまし 本稿では、特徴点追跡法について述べる。提案手法は、回転・スケール変化・照明変化による画像の変化に不変な SIFT(Scale Invariant Feature Transform) 特徴量を用い、Mean-Shift 探索により追跡を行う。評価実験により、回転やスケール変化を含む対象物体の移動に対しても追跡可能であることを確認した。また、追跡結果の表現法として、追跡点をセグメンテーションすることなく動線を表現する方法について述べる。点の移動方向と密度により表示する色を決めることで、移動の方向と頻度を表現する。本手法により、長時間に及ぶ人の追跡結果をビジュアライゼーションし、移動の流れが表現されていることを示す。

Mean Shift-based Point Feature Tracking using SIFT

Yuji Tsuduki† Hironobu Fujiyoshi† Takeo Kanade††

†Dept. of Computer Science, Chubu University

††The Robotics Institute, Carnegie Mellon University

E-mail:tsuduki@vision.cs.chubu.ac.jp, hf@cs.chubu.ac.jp, tk@cs.cmu.edu

Abstract This paper presents a method of point feature tracking using SIFT(Scale Invariant Feature Transform). Our approach uses the Mean-Shift searching to track a point based on the information obtained by SIFT. Since the SIFT feature is invariant to changes caused by the rotation, scaling, and illumination, we can obtain higher tracking performance than the conventional approach. Using the trajectory of the points obtained by the proposed method, it is possible to visualize the traffic line of pedestrians.

1 はじめに

Structure from Motion による 3 次元形状の復元 [1] や、移動物体のセグメンテーション [2] には長時間に及ぶ特徴点追跡が必要である。特徴点追跡の手法として既に KLT 法が [3] 提案され、多くのアプリケーションで用いられている。KLT 法は、局所領域における各点の動きは同一であると仮定し、弛緩法により目的関数を最小化する手法である。微小時

間における領域は平行移動のみしかしない、照明の変化による輝度値の変化がない、という状態を仮定して移動先を求める。そのため、対象物体の運動に回転やスケール変化を含む場合や、照明の変化による輝度値の変化が激しい場合、特徴点の追跡に失敗することがある。KLT 法を改良し、回転やスケール変化などにも対応可能な手法 [5],[6] も提案されているが、目的関数を決める際に運動モデルを仮定する必要があり、平行移動成分と回転・スケール変化

成分を別々に求める必要がある。

本稿では、このような対象とする運動を仮定する必要がなく、回転やスケール変化に対しても頑健な特徴点追跡手法について提案する。提案手法は、SIFT[7](Scale Invariant Feature Transform)により記述された特徴量を重みとして Mean-Shift 探索 [8],[9],[10] することで特徴点の追跡を行う。Mean-Shift による追跡は、局所的な重み分布に対する山登り計算であるため、高速な処理が可能である。特徴量にはカラーヒストグラムが用いられる場合が多く、ヒストグラム間類似度関数に従い求められる各画素の重みの分布に対して探索を行うため、追跡対象の形状変化や部分的なオクルージョン問題に対して頑健である。文献 [12] では色特徴以外に形状特徴として 3 方向のエッジを手がかりとすることで、さらに頑健な追跡を提案している。ただし、これらの特徴は領域を対象とした追跡に適したものであり、点特徴の追跡には適していない。提案手法で用いる SIFT 特徴は、各画素に回転・スケール変化・照明変化による画像の変化に不変な特徴量を記述することが可能であるため、Mean-Shift 探索による特徴点の追跡が可能となる。

次に、特徴点追跡結果の応用として、人の流れである動線のビジュアライゼーションについて述べる。各追跡点において、SIFT により検出されたスケールを考慮することで、移動体領域全体の動きを表現する事が可能である。

2 SIFT

SIFT は、特徴の抽出に適した点（以下、キーポイントと呼ぶ）を検出と、回転・スケール変化・照明変化に不変な特徴量を記述する 2 段階からなる。検出段階はスケール空間極値検出、キーポイントのローカライズからなり、記述段階はオリエンテーション割り当てと、SIFT Descriptor により特徴量の抽出からなる。

2.1 スケール空間極値検出

SIFT は、スケール変化による不変性を得るため、対象とする点の特徴をより表現できる近傍領域の範囲を自動的に決定する。範囲を決めるパラメータをスケールといい、キーポイント検出時に最適なスケールを決める。まず、スケール s のガウス関数 $G(x, y, \sigma)$ と入力画像 $I(x, y)$ を畳み込んだ平滑化画

像 $L(x, y, \sigma)$ を以下の式により作成する。

$$L(x, y, \sigma) = G(x, y, \sigma) * I(x, y) \quad (1)$$

$$G(x, y, \sigma) = \frac{1}{2\pi\sigma^2} \exp\left(\frac{-(x^2 + y^2)}{2\sigma^2}\right) \quad (2)$$

スケールの異なる平滑化画像間の差分処理を DOG(Difference of Gaussian) といい、 $D(x, y, \sigma)$ で表す。 $D(x, y, \sigma)$ は、以下の式より求める。

$$D_i(x, y, \sigma) = L(x, y, \sigma_{i+1}) - L(x, y, \sigma_i) \quad (3)$$

次に、図 1 に示すように DOG 画像 $D(x, y, \sigma)$ から極値の検出を行う。注目画素（図中 \times 印）が 26 近傍領域において極値（最大値もしくは最小値）である場合、キーポイント候補点とする。また、このとき極値となる DOG 画像のスケールをキーポイントの特徴抽出に用いる。

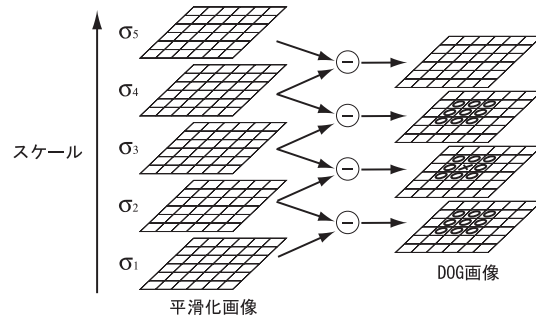


図 1: 極値検出の流れ

2.2 キーポイントのローカライズ

検出されたキーポイント候補点には、開口問題が発生しやすい点が含まれる。そこで、コントラストの低い点やエッジ上の候補点を削除し、より頑健な点をキーポイントとして選択する。

2.3 オリエンテーション割り当て

検出された各キーポイントに対して、オリエンテーションを割り当てる。キーポイントが検出された平滑化画像の各画素の勾配 $m(x, y)$ とその勾配方向 $\theta(x, y)$ を、以下の式より求める。

$$m(x, y) = \sqrt{f_x(x, y)^2 + f_y(x, y)^2} \quad (4)$$

$$\theta(x, y) = \tan^{-1} \frac{f_y(x, y)}{f_x(x, y)} \quad (5)$$

$$\begin{cases} f_x(x, y) = L(x+1, y) - L(x-1, y) \\ f_y(x, y) = L(x, y+1) - L(x, y-1) \end{cases} \quad (6)$$

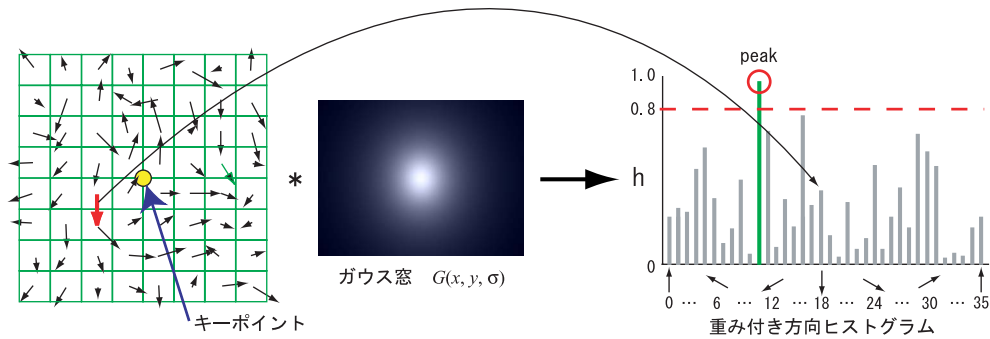


図 2: 重み付き方向ヒストグラム

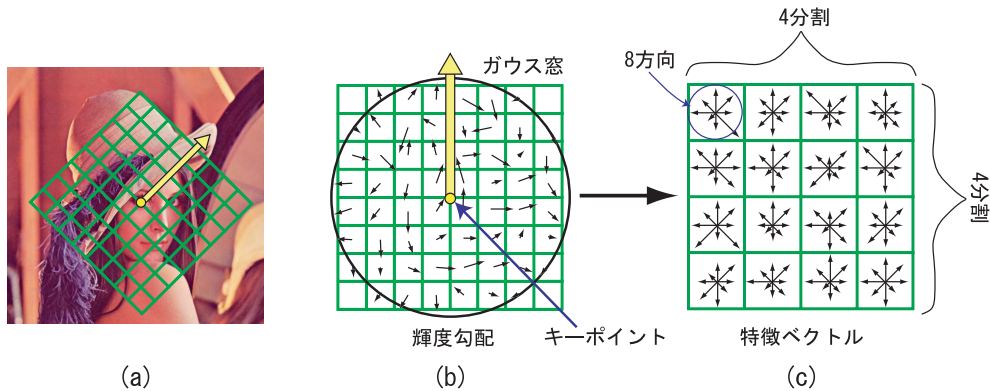


図 3: 特徴量記述

求めた勾配の大きさ m と勾配方向 θ から、36 方向に分割した重み付方向ヒストグラムを作成する (図 2 参照)。次に、作成した 36 方向のヒストグラムから、最大値の 80% 以上になる方向をキーポイントのオリエンテーションとして割り当てる。図 2 では、キーポイントに割り当てられるオリエンテーションは 1 方向のみであるが、80% 以上のピークが複数ある場合、それら全てがキーポイントのオリエンテーションとなる。例えば、キーポイントがコーナー上の点である場合、1 つのキーポイントに対して 2 方向のオリエンテーションが割り当てられることになる。

2.4 SIFT Descriptor による特徴量抽出

キーポイントの周辺領域を、図 3(a) に示すように割り当てられたオリエンテーション方向に回転する。このように、キーポイントの向きにあわせて特徴量を記述するため、回転に対する不変性が得られる。特徴量の記述の際には、図 3(b) に示すように、ガウス窓を用いて中心付近が強くなるように重み付けをする。このガウス窓の大きさは、キーポイントが検出された DOG 画像の平滑化スケールによって決定される。画像サイズが 2 倍であればスケールも

2 倍となり、構造的に同一の範囲における特徴量の記述を行うことになる。これにより、スケール変化に対する不変性が得られる。次に、分割した領域ごとに方向ヒストグラムを作成する。図 3(c) のように、領域を 4×4 の 16 に分割し、それぞれに 8 方向のヒストグラムを作成するため、128 次元のベクトルの特徴量を持つことになる。

3 SIFT 特徴を用いた Mean-Shift 追跡

本手法は、SIFT 特徴を用いた Mean-Shift アルゴリズムにより、キーポイントの移動先を決定する。本手法の追跡処理の流れを図 4 に示す。

3.1 初期追跡点検出

初期フレームの画像より初期追跡点を検出する。初期追跡点には、SIFT で検出されるキーポイントを用いる。各点には、SIFT で記述された 128 次元のベクトル $\mathbf{v} = (v_0, \dots, v_{127})$ とスケール s を参照用データとして保持する。

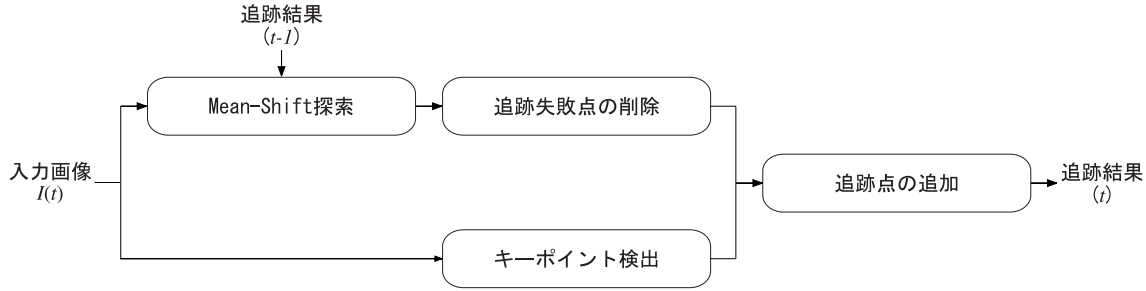


図 4: 本手法の流れ

3.2 Mean-Shift 探索

SIFT は、スケールによって特徴を抽出する範囲を変えるため、追跡対象の画像上での大きさが変化する場合は、適応してスケールを合わせる必要がある。そこで、提案手法の Mean-Shift 探索では、画像空間の移動量 $\Delta x = (\Delta x, \Delta y)$ とスケール空間の移動量 Δs を求める。SIFT 特徴量から求めた重み分布を用いて、画像空間とスケール空間の移動量を同時に求め、現フレームでの最も最適な位置とスケールに移動するよう計算を繰り返す。

step1 重み分布の計算

追跡中心 \mathbf{x} の周辺画素 $\mathbf{x}_i (i = 0, \dots, N)$ 及び参照スケール s の周辺スケール $s_j (j = 0, \dots, M)$ の SIFT 特徴ベクトルと \mathbf{x} における参照ベクトル \mathbf{v} との距離より重み $\omega(\mathbf{x}_i, s_j)$ を求める。

$$\omega(\mathbf{x}_i, s_j) = \exp\left(\frac{-d(\mathbf{x}_i, s_j)^2}{2\sigma_d^2}\right) \quad (7)$$

$$d(\mathbf{x}_i, s_j) = \|\text{SIFT}(\mathbf{x}_i, s_j) - \mathbf{v}\| = \sqrt{\sum_{k=0}^{127} (v_{i,k} - v_k)^2} \quad (8)$$

$$\text{SIFT}(\mathbf{x}_i, s_j) = \mathbf{v}_i = (v_{i,0}, \dots, v_{i,127}) \quad (9)$$

$\text{SIFT}(\mathbf{x}_i, s_j)$ は、座標 \mathbf{x}_i においてスケール s_j で SIFT 特徴ベクトルを出力する関数である。

step2 移動量の計算

求めた重み $\omega(\mathbf{x}_i, s_j)$ を用いて画像空間上の移動量 $\Delta x, \Delta s$ を求める。前フレームの追跡領域周辺に注目した探索を行うため、式 (12) に示す画像空間のカーネル関数 $K(\mathbf{x}, s, \sigma)$ を用い、移動量 $\Delta x, \Delta s$ を次式より求める。

$$\Delta \mathbf{x} = \frac{\sum_{i=0}^N \sum_{j=0}^M K(\mathbf{X}_i - \mathbf{X}, s_j - s, \sigma) \omega(\mathbf{X}_i, s_j) (\mathbf{X}_i - \mathbf{X}_0)}{\sum_{i=0}^N \sum_{j=0}^M |K(\mathbf{X}_i - \mathbf{X}, s_j - s, \sigma)| \omega(\mathbf{X}_i, s_j)} \quad (10)$$

$$\Delta s = \frac{\sum_{i=0}^N \sum_{j=0}^M K(\mathbf{X}_i - \mathbf{X}, s_j - s, \sigma) \omega(\mathbf{X}_i, s_j) (s_j - s)}{\sum_{i=0}^N \sum_{j=0}^M |K(\mathbf{X}_i - \mathbf{X}, s_j - s, \sigma)| \omega(\mathbf{X}_i, s_j)} \quad (11)$$

$$K(\mathbf{x}, s, \sigma) = \exp\left(\frac{-(x^2 + y^2 + s^2)}{2\sigma^2}\right) \quad (12)$$

移動量 Δx から、 $\mathbf{x}' = \mathbf{x} + \Delta x$ により、追跡点座標をシフトする。同様に、移動量 Δs から $s' = s + \Delta s$ により、追跡点のスケールをシフトする。

step3 繰り返し処理

これらの処理を $|\Delta x| < \epsilon_{xy}$ 、 $|\Delta s| < \epsilon_s$ を満たすまで Neab-Shift 探索を繰り返し、最適な追跡点座標とスケールを求める。 ϵ_{xy} と ϵ_s はしきい値である。

3.3 追跡失敗点の削除

移動後の座標 \mathbf{x} での SIFT の特徴ベクトルを求め、式 (8) により求めた参照ベクトルとの距離 $d(\mathbf{x}, s)$ がしきい値以上の場合、追跡に失敗したと判断し削除する。

3.4 追跡点の追加

新たに画像内に入れた移動体の追跡を行うために、特徴点の追加を行う。初期追跡点と同様に、キーポイントとして検出された点を新たな追跡点として登録する。ただし、追跡中の点と同じ位置が検出される場合もあるため、各追跡点から一定以上の距離があるキーポイントのみを追加する。

4 評価実験

合成画像を用いた実験により提案手法の有効性を示す。

4.1 実験方法

定量的に各手法の追跡精度を比較するため、合成画像による実験を行う。背景画像に追跡対象画像を貼り付けた画像を、追跡対象画像を 5 種類としたときの、平行移動、回転、拡大の動きを持つ合成画像を生成する。図 7(a) に合成した画像の例を示す。各

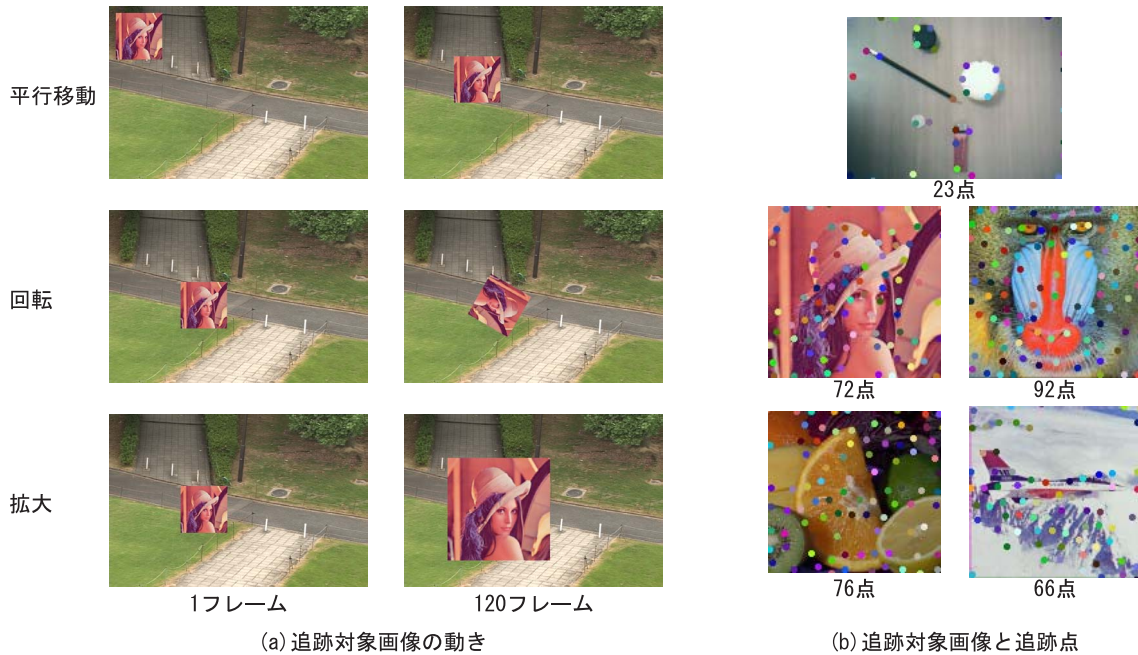


図 5: 実験用画像の例

シーケンスは 180 フレームである．提案手法と KLT 法により追跡を行い，追跡結果を比較する．本実験では追跡性能のみを評価するため，提案手法，KLT 法とも失敗点の削除は行わない．また，追跡対象の特徴点は KLT 法により求めたものとする．図 7(b) に各追跡対象画像における初期点とその数を示す．

4.2 実験結果

初期点と，実験用画像の合成に用いた変換行列より真値を求め，追跡実験結果との誤差（ユークリッド距離）を計算する．あるしきい値以下の誤差で追跡した点の割合を追跡成功率とし，特徴点追跡性能の評価に用いる．図 6 に追跡対象 5 種類のシーケンスでの全フレーム結果から求めた追跡成功率を示す．横軸がしきい値，縦軸が追跡成功率である．また，しきい値を 5 ピクセルとした際の追跡成功率を表 2 に示す．

表 1: しきい値 5 での追跡成功率 [%]

	平行移動	回転	拡大
提案手法	87.1	79.4	72.0
KLT 法	24.5	44.2	74.3

提案手法は，平行移動，回転において KLT 法よりも高い追跡成功率を得た．平行移動や回転では SIFT 特徴は不変であるため，頑健な追跡が可能である．また，拡大においては，本実験で用いた KLT 法で選択される追跡点はコーナー点が多く存在し，拡大に

よる変化の影響を受けにくいいため，KLT 法でも追跡が可能である．提案手法は，逐次，適応的にスケールを探索するため，拡大後も構造的に同じ範囲で特徴を抽出することができる．そのため，約 70% の点を追跡することができた．

4.3 非剛体の追跡例

提案手法は，アフィン変形等のモデルを仮定していないため，非剛体の動きに対する追跡が期待できる．図 7 に，屋外で撮影した人の動きに対する提案手法と KLT 法による追跡結果を示す．各点は 50 フレームの軌跡を表している．提案手法は，KLT 法に比べ，より多くの点を長時間にわたり追跡できていることが分かる．

5 動線のビジュアライゼーション

動線等の人の流れを表現する際，人領域のセグメンテーションを必要とする．しかし，人領域を正しくセグメンテーションすることは難しい問題である．本稿では，特徴点追跡の結果を用い，人の流れである動線のビジュアライゼーション手法について述べる．

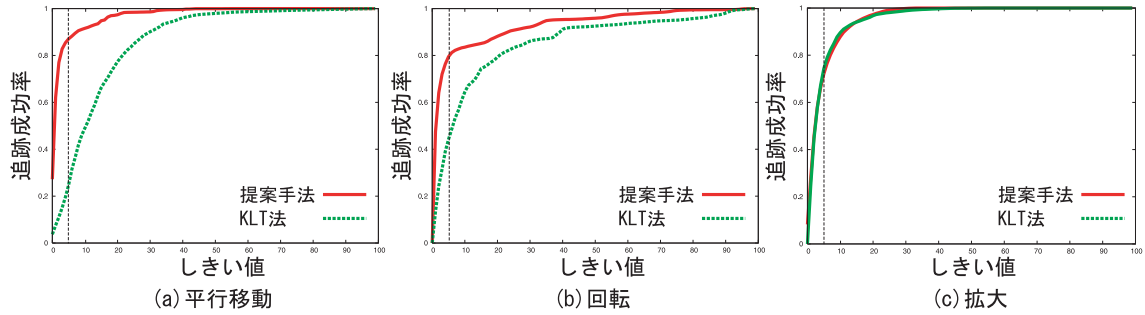


図 6: 実験結果

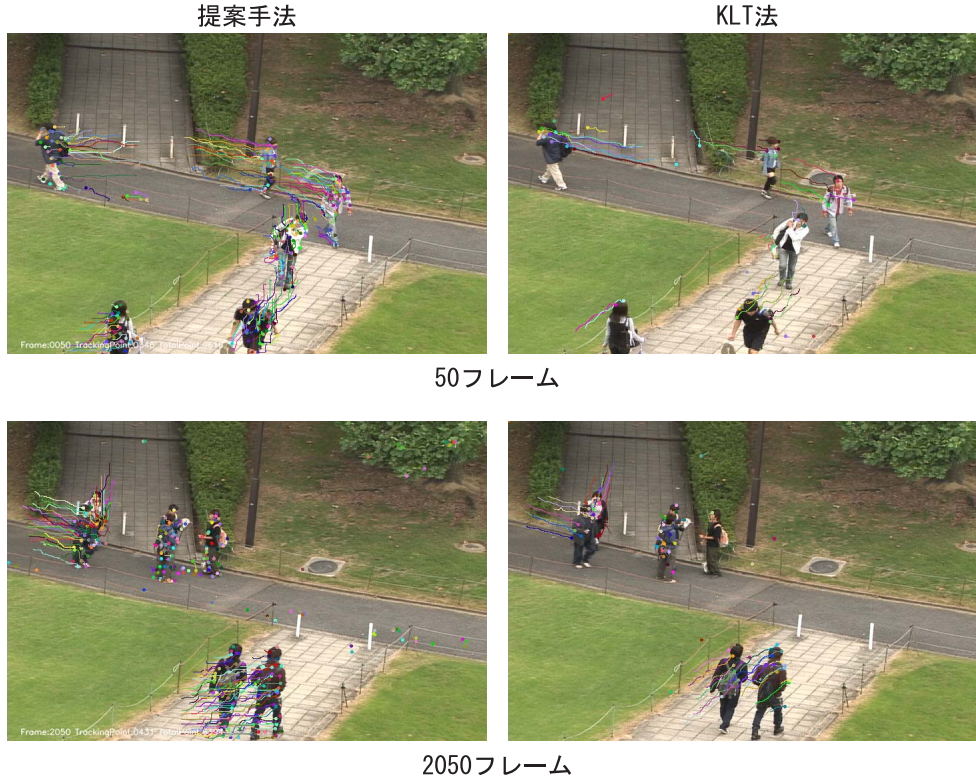


図 7: 人の追跡結果

5.1 ビジュアルライゼーション手法

動線のビジュアルライゼーションには、移動方向と頻度を読み取ることが重要である。しかし、人の特徴点追跡では、手や足の動き等の全体の動きとは異なる細かな動きが発生しやすい。このような点をビジュアルライゼーションに用いると、助長な情報が多くなり見づらい表示となる。そこで、フレーム間で移動方向に急激な変化がない点を用いるために、以下の基準により追跡点を選択する。

$$\cos \theta = \frac{\mathbf{v}_t \cdot \mathbf{v}_{t-1}}{|\mathbf{v}_t| |\mathbf{v}_{t-1}|} > th \quad (13)$$

$$\mathbf{v}_t = (\mathbf{x}_t, \mathbf{x}_{t-1}), \mathbf{v}_{t-1} = (\mathbf{x}_{t-1}, \mathbf{x}_{t-2}) \quad (14)$$

移動方向に大きな変動がなければ $\cos \theta$ は 1 に近い値となる。本手法では、 $Th = 0.9$ としたときの追跡

点を用いる。選択された追跡点の動きを表現するために、移動方向に対応した色相から色を決定する。移動方向 d は 8 方向とし、座標 \mathbf{x} における方向 d の色強度 $f_d(\mathbf{x})$ を以下の式より求める。

$$f_d(\mathbf{x}) = \sum_{t=1}^T \sum_{i=1}^N \delta(\mathbf{x} - \mathbf{x}_i^t, s_i) \quad (15)$$

T は総フレーム数、 N は追跡点数である。 \mathbf{x}_i^t はフレーム t における i 番目の追跡点の座標であり、 δ はガウス分布である。このときガウス分布の標準偏差として追跡点のスケール s_i を用いる。スケールは、テクスチャの少ない領域で大きくなる傾向がある。これは、テクスチャの少ない領域では、狭い範囲の画素から十分な特徴を得られないためである。テクスチャが少ない領域では、追跡点の密度が低く

なるが、定数倍したスケールを用いてガウス分布を決定する事で、密度の低さを補う。図 8 に示すように、追跡点の密度が高い領域は移動方向に対応した色が強く表現されることになる。

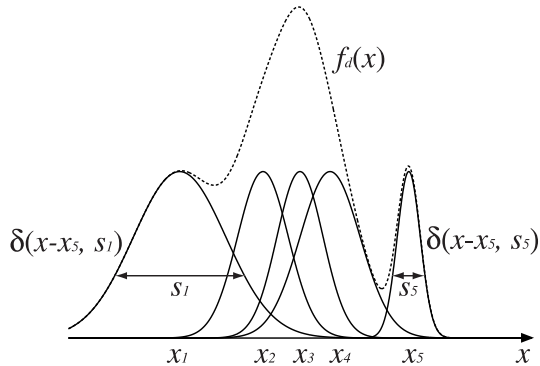


図 8: 点の密度による色の強さ

図 9 に、あるフレームにおける追跡点を用いた、人のビジュアライゼーションの例を示す。図 9(b) における追跡点の大きさはスケールを表している。

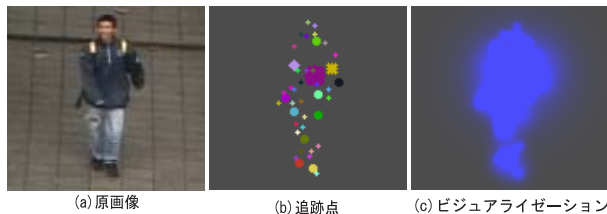


図 9: スケールを考慮した表現

このように、移動体領域をセグメンテーションしなくても、特徴点とそのスケールを用いることで、対象物体の大まかな領域を表現することができる。

5.2 結果

図 10(a) に、約 1 時間 (100,000 フレーム) の追跡結果から動線をビジュアライゼーションした結果を示す。図 10(a) おける左下の円は移動方向の色を示している。この例では、図中 A の領域では左に横切る人が多く、B の領域では移動方向が左右で反対となっていることがわかる。このように、長時間に亘る特徴点追跡の結果をビジュアライゼーションすることで、定常的な流れを確認することが可能となる。また、動線表示の対象時間を、約 2 秒間隔 (60 フレーム) とした際のビジュアライゼーション結果を図 10(b)(c)(d) に示す。短時間における表示では、図 10(b)(d) の A の領域とは異なる右方向へ

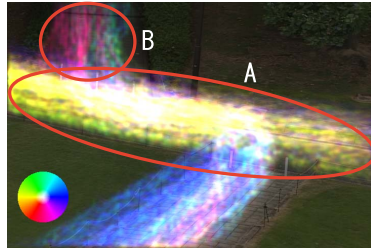
の移動が確認できる。従って、移動体領域において、図 10(a) の画像と図 10(b)(c)(d) の各画像との差分をとることで、定常的な流れとは異なる動きの自動検出が期待できる。

6 おわりに

SIFT 特徴を用いた Mean-Shift による特徴点追跡手法について提案した。提案手法は SIFT 特徴の頑健性により、平行移動、回転に対して精度良く追跡できることを確認した。追跡対象が拡大する場合においても、SIFT のスケールを適応的に更新する事で、追跡が可能である。また、人の追跡を例に、特徴点追跡結果のビジュアライゼーション法を示した。点の密度とスケールにより色の強度を変えることで、セグメンテーションを必要とせず、人の流れを表現可能であることを確認した。今後は、定常的な流れとは異なる動きの自動検出に取り組む予定である。

参考文献

- [1] C. Tomasi and T. Kanade, "Shape and Motion from Image Streams under Orthography: A Factorization Method", Int. Journal of Computer Vision, Vol. 9, No. 2, pp. 137-154,1992.
- [2] J. P. Costeira and T. Kanade, "A multibody factorization method for independently moving objects", Int. Journal of Computer Vision, Vol. 29, No. 3, pp. 159-179,1998.
- [3] C. Tomasi and T. Kanade, "Detection and tracking of point features", Technical report, CMU-CS-91-132, 1991.
- [4] J. Shi and C. Tomasi, "Good Features to Track",1994 IEEE Conference on Computer Vision and Pattern Recognition (CVPR'94), pp. 593-600, 1994.
- [5] 深尾隆則, 金出武雄, "2 段階特徴点追従アルゴリズム", 情報処理学会研究報告, 2003-CVIM-141, pp. 103-110.
- [6] 深尾隆則, 金出武雄, "オクルージョンに対してロバストな 2 段階特徴点追従アルゴリズム", 情報処理学会研究報告, 2003-CVIM-141, pp. 111-118.
- [7] David G. Lowe, "Distinctive image features from scale-invariant keypoints", Journal of Computer Vision, 60, 2, pp. 91-110, 2004.
- [8] D. Comaniciu and P.Meer, "Mean shift analysis and applications", in IEEE Int'l Conf. Comp. Vis, 1999, pp. 1197-1203.
- [9] D. Comaniciu, V. Ramesh and P.Meer, "Real-time tracking of non-rigid objects using mean



(a) 約1時間の追跡結果



(b) 約2秒間の追跡結果 (50フレーム)



(c) 約2秒間の追跡結果 (1050フレーム)



(d) 約2秒間の追跡結果 (2050フレーム)

図 10: ビジュアライゼーション結果

shift”, in IEEE Computer Vision and Pattern Recognition, pp.II:142-149, 2000.

- [10] 川中 大樹, 中島 平, 岡谷 貴之, 出口 光一郎, “平均値シフト法を用いた複数物体追跡の研究”, IPSJ SIG Technical Report CVIM No.138(2), 2003.
- [11] 末岡 千枝, 市川 清人, 三田 雄志, 堀 修, “Mean Shiftを用いたリアルタイム顔画像追跡手法に適した特徴量の提案”, 画像の認識・理解シンポジウム (MIRU 2006) 論文集, pp.376-381, 2006.
- [12] K. She, G. Bebis, H. Gu and R. Miller, “Vehicle Tracking Using On-Line Fusion of Color and Shape Features”, in IEEE Conference on Intelligent Transportation Systems, 2004.