Vol.2012-CVIM-180 No.58
2012/1/20

Kinect を用いた顔・髪の毛の追跡

鈴 木 一 正[†1] 戚 意 強[†1] 呉 海 元[†1]

本研究グループが提案している K-means Tracker では，追跡対象と類似している色を持つ背景画素が追跡領域に混入した場合，その背景画素が誤ってターゲット画素としてクラスタリングされ，追跡が不安定になるという問題がある．この問題を解決するために，追跡ターゲットとカメラ間の 3 次元空間の奥行き情報を追加した拡張 K-means Tracker を提案している．本論文では，拡張 K-means Tracker の色情報と距離情報を Kinect で獲得する手法を提案し，この手法を用いた顔や髪の毛のビデオレートでの追跡システムを提案する．また，顔と髪の毛を同時に追跡することで，顔の向きに関わらず安定な追跡が行える．

Face and Hair Tracking Using Kinect

Kazumasa Suzuki,[†1] YiQiang Qi[†1] and Haiyuan Wu[†1]

We have proposed an object-tracking algorithm called as EK-means tracker, which has two improvements are made to the conventional K-means tracker to solve the instability problem that occurs when some background objects show similar colors to the target or the size of the target object changes significantly. The first one is introducing of the depth information as the sixth feature into the original 5D feature space for describing pixels. In this paper, we propose a face and hair tracking system at video-rate by using Kinect to capture color images and depth images. Our system can perform stable tracking regardless of face direction because it tracks the face and hair simultaneously.

1. K-means Tracker[1)] の概要と問題点

K-means Tracker は，追跡対象と背景の両方に複数のクラスタ中心を割り当て，K-means クラスタリングによってサーチエリア内の各画素をターゲットと非ターゲットに分けることにより非剛体物体を追跡する手法である．追跡対象を構成する画素の色の類似性と空間的近接性を同時に表現するために，各画素の特徴を 3 次元色空間 $\mathbf{c} = [\mathbf{Y}\,\mathbf{U}\,\mathbf{V}]^{\mathrm{T}}$ と 2 次元座標空間 $\mathbf{p} = [\mathbf{x}\,\mathbf{y}]^{\mathrm{T}}$ からなる 5 次元ベクトル $\mathbf{f} = [\mathbf{c}\,\mathbf{p}]^{\mathrm{T}}$ として扱う．2 つの画素 a, b 間の距離 $d(\mathbf{f_a}, \mathbf{f_b})$ は，色空間内のユークリッド距離と画像座標空間内のユークリッド距離を重み α により次式で定義している．

$$d(\mathbf{f_a}, \mathbf{f_b}) = ||\mathbf{c_a} - \mathbf{c_b}||^{\mathbf{2}} + \alpha||\mathbf{p_a} - \mathbf{p_b}||^{\mathbf{2}} \tag{1}$$

また，クラスタが未知である注目画素 $\mathbf{f_u}$ からターゲットクラスタ中心 $\mathbf{f_T}$ への最短距離 D_T，非ターゲットクラスタ中心 $\mathbf{f_{NT}}$ への最短距離 D_{NT} は，以下のように定義される．

$$D_T = \min_{i=1\sim n}\{d(\mathbf{f_{Ti}}, \mathbf{f_u})\} \tag{2}$$

$$D_{NT} = \min_{j=1\sim m}\{d(\mathbf{f_{NT_j}}, \mathbf{f_u})\} \tag{3}$$

ここで m, n は，それぞれターゲットクラスタ中心と非ターゲットクラスタ中心の個数である．D_T と D_{NT} を比較することによって，注目画素 $\mathbf{f_u}$ がターゲットクラスタか非ターゲットクラスタにクラスタリングされる．ターゲットとしてクラスタリングされた画素集合に可変楕円を当てはめ，楕円のパラメータ (長軸，短軸，傾き，重心位置)，ターゲットクラスタ中心，非ターゲットクラスタ中心を毎フレーム更新することで，追跡対象の大きさや形状が変化しても安定な追跡が行える．

しかし，K-means Tracker には次のような問題点がある．K-menas Tracker において画素間の距離を求める際に，背景画素がターゲットと類似色の場合は $||\mathbf{c_T} - \mathbf{c_u}||^{\mathbf{2}} \approx \mathbf{0}$ となり，また，ターゲットに近づくと $||\mathbf{p_T} - \mathbf{p_u}||^{\mathbf{2}} \approx \mathbf{0}$ となり，その結果 $d(\mathbf{f_T}, \mathbf{f_u}) \approx \mathbf{0}$ となる．

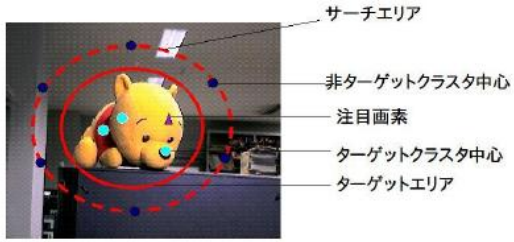

図 1 K-means Tracker の構成要素

†1 和歌山大学
Wakayama University

ⓒ 2012 Information Processing Society of Japan

つまり，追跡対象と類似している色を持つ背景画素がサーチエリアに混入した場合，5 次元特徴空間ではその背景画素を誤ってターゲット画素としてクラスタリングされる可能性が高く，その影響によりターゲット領域やサーチエリアが不安定となり，追跡が失敗してしまう場合がある (図 6，7 参照)．また，空間的距離を画像上のピクセル単位で計算しているため，ターゲットが小さいく写っているときは 2 次元座標空間内の距離は小さくなり，ターゲットが大きく写っているときは 2 次元座標空間内の距離は大きくなる．このように，画像内のターゲットの大きさによって，5 次元特徴空間内における 2 次元位置特徴ベクトルの割合が変わってくるという問題がある．

2. K-menas クラスタリング用の特徴空間の拡張

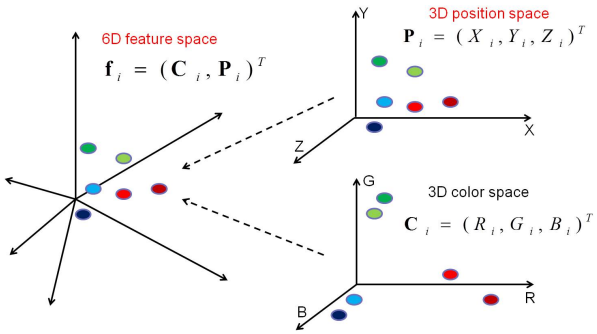

図 2 6 次元拡張特徴空間

K-means Tracker の類似背景問題を解決するために，K-means Tracker で定義した特徴空間を拡張する．3 次元ワールド座標系においてターゲットと背景が空間的に離れていると仮定すると，カメラからターゲット，背景までの 3 次元空間内の奥行き距離を利用することで，サーチエリア内に混入したターゲットと似たような色を持った背景画素を正しくクラスタリングすることができる．拡張 K-means Tracker では，従来の 5 次元特徴空間に奥行き特徴を新しく追加し，6 次元に拡張する（図 2 参照）．そこで，画素の位置を表現するベクトルを 3 次元位置空間 $\mathbf{p}$ に変更する．

$$\mathbf{p} = [\mathbf{X}\,\mathbf{Y}\,\mathbf{Z}]^{\mathrm{T}} \tag{4}$$

画素間の距離は，変更された $\mathbf{p}$ と式 (1) により求められる．このように，各画素をワールド座標系に変換した上で空間的近接性を求めることにより，類似背景の問題と画像内ターゲットサイズの変化による問題を同時に解決することができる．

3. 顔・髪の毛追跡システム

本論文では，Kinect で得られる色情報と距離情報を活用し，ビデオレートで安定な顔・髪の毛の検出・追跡システムを提案する．

3.1 Kinect で得られる距離画像

拡張 K-means Tracker で用いる距離情報は，これまでステレオカメラで取得していたが，ステレオ処理のコストが高く，サーチエリア内の全ての画素において距離を求めることは難しかった．本システムでは，入力画像（640x480 画素）の距離情報が，Kinect を用いることで高速（30fps）に取得できる．図 3 は，Kinect により得られた距離画像例である．明るいほど距離が近いことを意味している．Kinect から得られる距離画像はオクルージョンなどによって距離情報が得られない画素が多々ある．ここでは，距離情報が得られない画素については，その画素のエピポーラライン（水平方向）上において，左右それぞれの最近傍の距離情報がある画素の内，距離の遠い方がその画素の距離情報となるように補完を行なっている．

3.2 検出・追跡システム

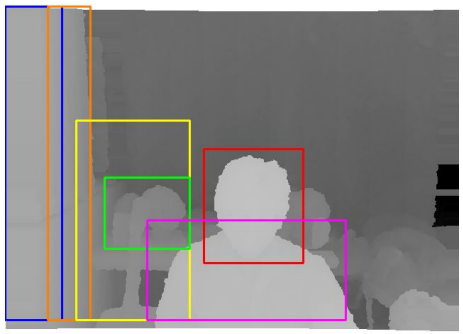

図 3 距離画像

本システムでは，顔や髪の毛といった追跡対象のクラスタ中心情報を得るために，まず，距離情報を用いた高速な顔検出法[2] を利用して顔検出を行う．この検出手法は，距離情報と

ⓒ 2012 Information Processing Society of Japan

実際の人物顔の大まかなサイズから画像上での顔サイズを推定することで，限られた位置とスケールのみの探索を行う．距離に応じた大きさだけの顔を探索し，識別回数を削減することによって，ピラミッドスキャンする場合よりも高速に顔を検出することができる．また，識別回数が大幅に削減されるため，誤検出を減少させる効果も得られる．図 3 の矩形は，同じような距離を持った画素の領域を表しており，各領域内を特定のスケールのみで探索する．探索を行う際の識別器には，OpenCV の顔検出ライブラリ関数[3),4)] を用いている．顔検出に成功した場合，顔あるいは髪の毛の追跡に移行する．

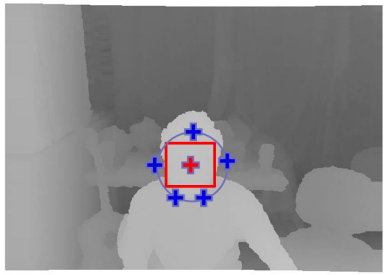

図 4　顔追跡の初期化

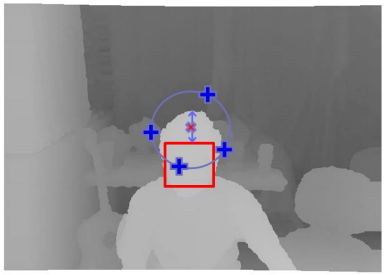

図 5　髪の毛追跡の初期化

追跡開始時は，顔検出によって得られた矩形を用いて初期のターゲット，非ターゲットクラスタ中心を設定し，追跡の初期化を行う．顔を追跡する場合は，顔検出によって得られた矩形の中心をターゲットクラスタ中心，矩形の外接円上に非ターゲットクラスタ中心を配置し，その円を初期フレームにおける可変楕円として追跡を開始する（図 4 参照）．髪の毛を追跡する場合は，検出矩形の上辺からデプスのエッジまでの中点をターゲットクラスタ中心，そこから顔サイズ分だけ離れた円周上に非ターゲットクラスタ中心を配置する（図 5 参照）．追跡時は，ターゲットとしてクラスタリングされた画素が少なくなった場合，対象の追跡に失敗したと判断し，再び検出に移行する．

4. 実　　験

4.1 類似色背景の比較実験

図 6，7 は，それぞれ同じ動画シーケンスに対して従来手法（K-means Tracker）と提案手法による追跡を行った結果を比較したものである．背景がターゲットと類似色の場合，従来の K-means Tracker では背景がターゲット画素としてクラスタリングされてしまい，対象領域が広がって追跡が失敗しているのに対し，距離情報を利用した提案手法では安定した追跡が行えていることがわかる．

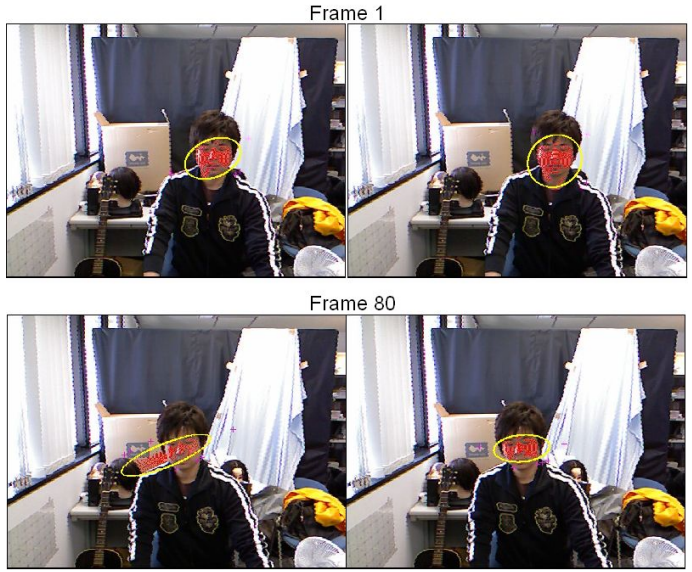

図 6　類似背景（ダンボール）の顔追跡結果例，(左) 従来手法，(右) 提案手法

図 7　類似背景（黒幕）の髪の毛追跡結果例，(左) 従来手法，(右) 提案手法

ⓒ 2012 Information Processing Society of Japan

4.2 顔・髪の毛の同時追跡

図 8，9 は，提案手法を用いて，顔だけを追跡した場合と，顔・髪の毛を同時に追跡した場合の比較結果である．顔だけの追跡では，追跡中に顔（肌色）領域が小さくなると図 8 のように追跡が失敗してしまうなど，顔の向きによって追跡が不安定になる場合がある．しかし，図 9 のように髪の毛も同時に追跡することで，顔がカメラに対して後ろを向いた場合でも追跡を継続することができる．

5. おわりに

本論文では，Kinect を用いることで，容易に高速・高精度な距離情報を得られることを利用し，K-menas クラスタリング用の特徴空間を 5 次元から 6 次元に拡張することによって従来の K-means Tracker の類似色背景問題と画像内のターゲットサイズ問題を解決できる手法を提案した．また，顔検出と顔・髪の毛追跡システムを構築し，ターゲットとの類似色が背景に混入した場合でも，ターゲットと背景の奥行きの違いにより安定な追跡を行うことができることを示し，提案手法の有効性を確認した．また，実験では 1 人の追跡対象について比較実験を行ったが，この提案手法は，複数対象を追跡することも可能である．

謝　　辞

本研究の一部は，文部科学省科学研究費補助金基盤研究 (c)(一般)21500171 の補助を受けている．

参 考 文 献

1) Chunsheng Hua, Haiyuan Wu, Qian Chen, and Toshikazu Wada. K-means tracker: A General Algorithm for Tracking People. Journal of Multimedia, Vol. 1 Issue.4, pp. 46-53, July 2006.
2) Haiyuan Wu, Kazumasa Suzuki, Toshikazu Wada, Qian Chen. Accelerating Face Detection by Using Depth Information. In Proceedings of the 3rd Pacific Rim Symposium on Advances in Image and Video Technology, pp 657-667, Jan 2009.
3) Paul Viola and Michael Jones, Robust Real-Time Face Detection, IJCV, Vol.57, No.2, pp.137-154, 2004.
4) OpenCV, http://www.intel.com/technology/computing/opencv
5) Yiqiang Qi, HaiyuanWu, A Pixel-wise tracking algorithm using stereo camera, The 7th Int. Conf. on Ubiquitous Robots and Ambient Intelligence (URAI 2010), pp.198-201, 2010

図 8　顔だけの追跡

図 9　顔・髪の毛の同時追跡

ⓒ 2012 Information Processing Society of Japan