

肩載せ PTZ カメラを用いた写真撮影システム

菅佐原 俊介[†] 河野 恭之[†]

関西学院大学大学院理工学研究科[†]

1. はじめに

本研究は、ユーザがハンドフレームを用いて捉えた景色・構図を肩載せカメラによって撮影するシステムを提案する。ハンドフレームの例を図1に示す。ハンドフレームとは、撮影者が写真の構図を検討するにあたって一般的に使う、両手で四角い囲いを作る動作のことを指す。ハンドフレームを用いた構図検討の手法は撮影者にとって容易かつ直感的である。しかし、決定した構図をカメラで撮影するには二つの問題がある。一つ目は、肉眼で捉えた風景とカメラレンズ越しの風景とが同一ではない点である。二つ目は、ハンドフレームから撮影の姿勢へ移行するまでに多くの動作を要する点である。そこで本研究では、ユーザがハンドフレームを用いて捉えた撮影領域と同等の構図を肩載せ PTZ (パン・チルト・ズーム) カメラで静止画撮影するウェアラブルなシステムの作成を目指す。淵ら[1]は、ユーザがハンドジェスチャで指定した範囲を天井に設置した3台のカメラを用いて撮影するシステムを開発している。このシステムはカメラを事前に配置する必要があるため、使用する環境が限定される。また撮影画像はユーザ視点の風景ではない。酒田ら[2]は、肩載せカメラとレーザポイントによって遠隔の協調作業を支援するシステムを提案している。作業者の動作を推定し、それに合わせてカメラのパン・チルトを制御することで常に定点の撮影・指示を行う手法が実装されている。酒田らの研究はカメラのズーム機能を使用していない。本研究では、カメラをズームさせた後も特定の領域に留まるために、ユーザの身体の揺れに対してロバストなパン・チルトの制御を行う。



図 1. ハンドフレーム

本システムの構成を図2に示す。ユーザは撮影写真の構図を確認するために HMD (ヘッドマウントディスプレイ) を装着し、頭部にカメラを装着する。また、撮影用の PTZ カメラを右肩に装着する。視界を確保するためのカメラとは別に撮影用のカメラを装着することで、高画質の写真を撮影することが可能になる。

本システムの処理の流れを図3に示す。まず頭部装着カメラの画像からユーザの手領域を抽出する。得られた手領域からハンドジェスチャを検出し、撮影領域の範囲の決定や撮影領域画像の取得を行う。撮影領域画像とは、撮影領域のみを抽出し矩形化した画像である。次に、撮影領域画像と同等の領域を肩載せカメラの画像から探索し、発見した領域を目標領域と決定する。目標領域が全画面に表示されるように肩載せカメラの PTZ を制御する。

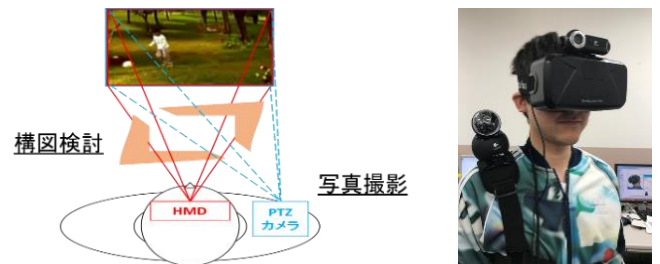


図 2. システム構成

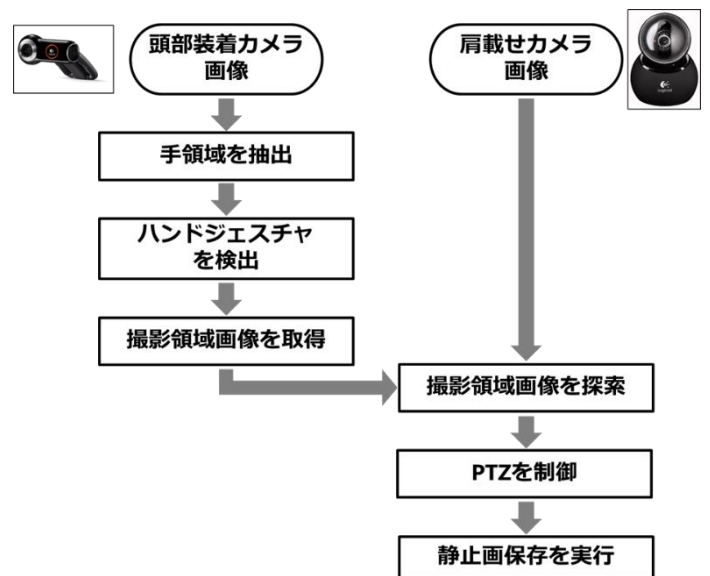


図 3. システムの流れ

2. 撮影領域画像の取得

本章では、頭部装着カメラの画像からユーザの手領域を抽出し、得られた手領域に対してジェスチャ判定処理を行い、システムがユーザの指定する撮影領域画像を取得するまでの過程を述べる。処理の流れを図4に示す。まず頭部装着カメラの画像から肌色領域を抽出するため、Sherrah

ら[3]が提案した肌色の色相範囲を設定し、肌色領域とそれ以外に二値化処理する。類似した色相の手領域と背景領域が重なる場合が考えられるため、彩度と明度の画像にソーベルフィルタを適用して得られたエッジを境界線とし、これらを分離する。分離後の画像から面積が最大二つの領域を抽出し、左下に位置するものを左手、右上に位置するものを右手の領域と決定する。得られた手領域抽出画像から、ハンドジェスチャの判定と認識を行う。本システムで認識するハンドジェスチャは、撮影領域指定、撮影待機、撮影実行である。ユーザは、本システムを利用するにあたってこれらのハンドジェスチャを順に行う。撮影実行ハンドジェスチャを認識した際、その時点で指定されている撮影領域内の画像を撮影領域画像として取得する。

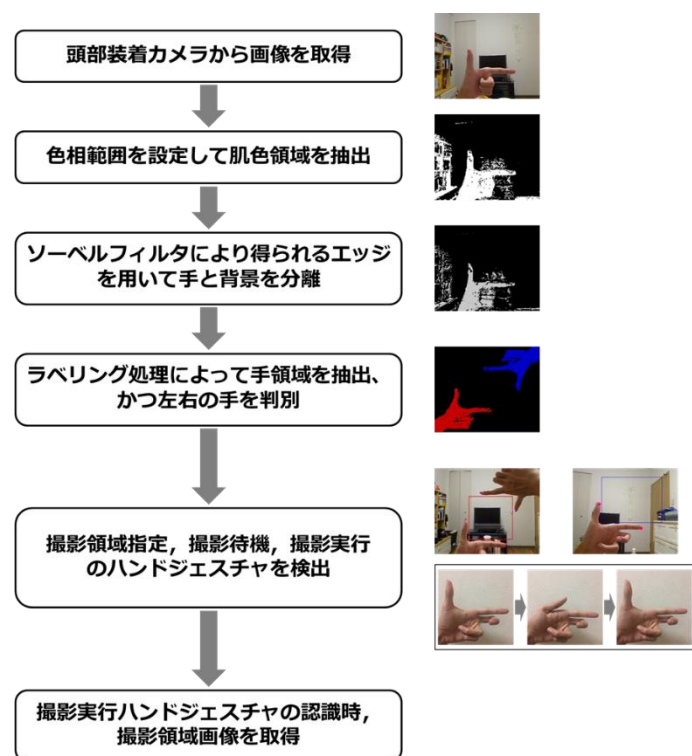


図 4. 撮影領域画像取得の流れ

撮影領域指定ハンドジェスチャは、ユーザがハンドフレームを用いて示す構図を撮影領域と決定する。システムがこのハンドジェスチャを認識している様子を図 5 に示す。左手と右手の指の付け根を対角とする長方形領域を撮影領域とし、赤い枠の描画で表している。ユーザは手の位置を動かすことで、撮影領域のサイズや位置を自由に変更することができる。撮影領域指定ハンドジェスチャを認識するために、ハンドフレームの検出を行う。ハンドフレームは、左右の手がそれぞれ適当な位置関係にあり、親指と人差し指が直角に開いている状態とする。まず親指と人差し指の指先を検出する。次に親指先を中心に、人差し指側へ円を描くように手領域の画素を探索（円探索）し、発見した点を付け根と仮定する。親指先と付け根、付け根と人差し指先を結ぶ各線の交差角度を求める。発見した点を中心に同様の探索を行い、そこで発見した点を用いて交差角度を求める。これらの処理を交差角度が増加し始めるまで続

ける。終了時までの最小の交差角度が直角に近似しているならば、その交差を作る点を付け根と確定し、手領域がハンドフレームであるとする。

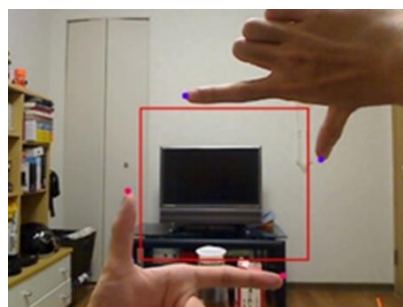


図 5. 撮影領域指定ハンドジェスチャの認識

撮影待機ハンドジェスチャは、ハンドフレームを検出している状態から、左手の人差し指の下に手領域を検出した際に認識する。これは、中指を伸ばす、握り拳を緩める等の少ない手の動きで実現できるため、使用が容易であり、撮影領域指定ハンドジェスチャからスムーズな移行が可能である。ハンドジェスチャの認識後は、撮影領域サイズを記録・固定し、右手の検出を停止する。また人差し指先に対する付け根の相対的な位置を記録する。これにより、ユーザが左手のみでシステムを扱えるようにして、右手が肩載せカメラの映像内に入り撮影のオクルージョンとなることを防ぐ。

撮影実行ハンドジェスチャは、親指を人差し指側に一定まで傾けて再び直角に立てる動作を検出した際に認識する。撮影実行ハンドジェスチャを図 6 に示す。親指の傾斜角度を計測するにあたり、親指先と人差し指先、記録済みの相対的な位置にある仮想の付け根との三点の角度を求める。仮想の付け根を用いることで、親指が傾いた際に、ハンドフレーム検出時の付け根の位置がずれる状況や、傾斜角度の計算が困難になる状況を防ぐことができる。撮影領域画像を取得して、肩載せカメラの PTZ 制御を開始する。

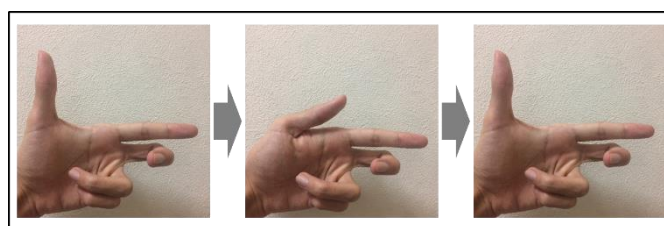


図 6. 撮影実行ハンドジェスチャ

3. PTZ 制御手法

本章では、ユーザが指定した撮影領域画像と同等の領域を肩載せカメラの画像中から探索し、発見した領域が画面全体に表示されるように肩載せカメラの PTZ を制御する手法を述べる。撮影する領域はユーザの身体の揺れ等に影響されて移動してしまうことがあり、これを追跡するため、特徴点の移動量に着目したロバストな制御を行う。本章で記す処理の流れを図 7 に示す。まず、発見した領域を目標

領域と決定し、目標領域がカメラ画像の中央に移動するようにカメラをパン・チルトする。このとき、1度パン・チルトしたときの、カメラ画像内の特徴点の移動量を推定する。目標領域が中央にあれば、目標領域がカメラ画像の全画面に表示されるようにズームを行う。次に、カメラ画像と目標領域の位置ずれを検出する。位置ずれが解消されるようにパン・チルトし、位置ずれが閾値以下であれば、カメラ画像を静止画保存する。

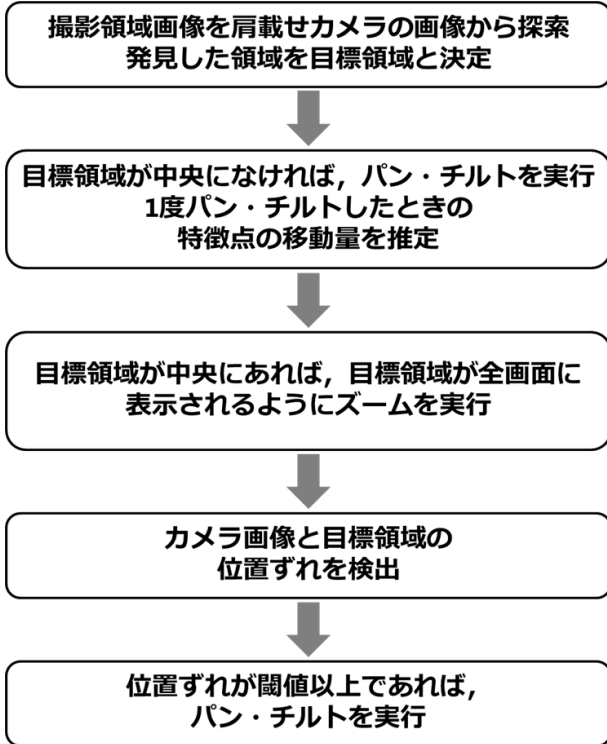


図 7. PTZ 制御の処理の流れ

3.1. 特徴点の移動量の推定

目標領域とカメラ画像との位置ずれを解消するために、肩載せカメラを1度パン・チルトしたときの、カメラ画像内の特徴点のx方向とy方向の移動量を把握する必要がある。特徴点の移動量は、パン・チルトを行う直前と直後のフレームを比較して特徴点のトラッキングを行うことで求める。観測を行う例を図8に示す。位置・形状が不変の被写体Aを捉えるとき、カメラを右へパンすれば、カメラ画像に映る被写体Aは左方向へ移動する。被写体Aの座標の差から、パンを行うことによって発生する移動量が求められる。ただし、特徴点の移動量は実空間上におけるユーザと撮影対象の位置関係に影響を受けるため、使用環境を限定しない本システムでは、写真撮影のためのPTZ制御を行いながら、その場における実際の移動量を観測する。特徴点の抽出にはAGAST特徴量[4]を、トラッキングにはLucas-Kanade法[5]を使用する。AGAST(Adaptive and Generic Accelerated Segment Test)は、画像中のコーナーを検出し、特徴点の抽出処理を高速で実行するアルゴリズムである。よって、AGAST特徴量を用いた特徴点抽出は、ウェアラブルシステムにおけるリアルタイムの処理を必要とする本研究に適している。Lucas-Kanade法では、

トラッキングすべき特徴点のサブセットをあらかじめ指定することで、高速なトラッキング処理を実行することができる。これも同様に、リアルタイムの処理を必要とする本研究に適している。

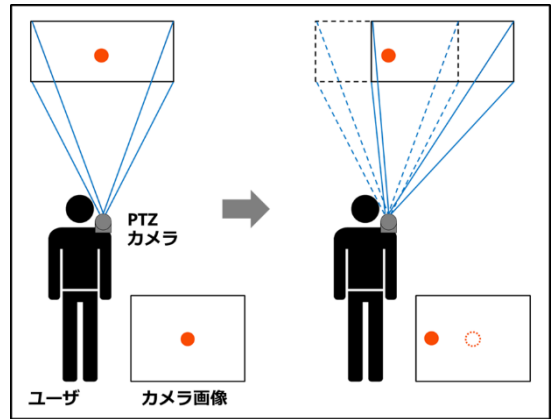


図 8. 特徴点の移動量の観測例

観測した移動量は、ユーザの身体の揺れや、検出する特徴点群の変化といったノイズに影響を受けて、誤差を含んでいると考えられる。そこで本研究では、カルマン(Kalman)フィルタ[5]を用いて誤差を含まない移動量を観測値から推定する手法を実装する。カルマンフィルタとは、誤差を含む観測値からある線形動的システムの適切な状態を推定、制御するための処理である。過去に学習した情報を用いてシステムの次の状態を推定する予測フェーズと、現在の観測値を取得し推定法のモデルの調整する修正フェーズを、時刻のワンステップ毎に実行する。それぞれの時刻における予測値と観測値を比較して推定法を改良していくので、データの蓄積に従って予測値から誤差を減らすことができる。本研究では、パン実行時とチルト実行時とで別に移動量を推定するため、二つのカルマンフィルタを作成する。パンまたはチルトを実行する前に移動量を予測し、実行後は現実の移動量の取得と、推定法のモデルの調整を行う。

3.2. 目標領域の追跡

目標領域はユーザの身体の揺れに影響されてカメラ画像内で移動することがあるため、これを追跡するためのPTZ制御を適宜行う。ズームを行った後、目標領域全体をカメラ画像内に収めたとき、カメラ画像を静止画保存する。本システムで行うPTZ制御は、二つの段階に分類される。二つのPTZ制御の概要を記したものを表1に示す。第一段階の目標領域に「接近する」ためのPTZ制御は、肩載せカメラの動作が初期の状態から、パン・チルトすることによって目標領域を中央に収め、ズーム倍率を調整するまでの処理である。カメラ画像から目標領域を追跡するために、テンプレートマッチングを行う。第二段階の目標領域に「留まる」ためのPTZ制御は、ズームを行った後、カメラ画像と目標領域との位置ずれを解消するためにパン・チルトを制御する処理である。カメラ画像から目標領域を追跡するために、特徴点トラッキングを行う。

表 1 PTZ 制御の概要表

	第一段階： 「接近する」ための PTZ 制御	第二段階： 「留まる」ための PTZ 制御
目標領域の追跡手法	テンプレートマッチング	特徴点トラッキング
パン・チルト	目標領域を中央に移動させるように実行	目標領域との位置ずれから必要な回転角度を計算し実行
開始条件	撮影実行ハンドジェスチャを検出	ズームの実行を確認
終了条件	カメラ取得画像の中央付近に目標領域を検出	目標領域との位置ずれが閾値以下
終了時の処理	目標領域画像を取得 ズームを実行	カメラ画像の静止画保存を実行

頭部装着カメラの画像からユーザの撮影実行ハンドジェスチャを検出した際、撮影領域画像を取得し、第一段階の「接近する」ための PTZ 制御を開始する。まず、肩載せカメラの画像から目標領域を探索する。探索手法に、撮影領域画像をテンプレート画像とした、テンプレートマッチングを用いる。図 9 は、頭部装着カメラ画像中の撮影領域を青色の枠で、テンプレートマッチングで発見した肩載せカメラ画像中の目標領域を緑色の枠で示している。発見した目標領域の座標を取得し、目標領域が中央になれば、中央に移動するようにカメラのパン・チルトを制御する。このとき、小刻みにパン・チルトを行い、3.1 節で述べた特徴点の移動量の観測と推定を行う。目標領域をカメラ画像の中央に検出したとき、目標領域がカメラ画像の全画面に表示されるようにカメラのズームを行う。ズーム倍率は、カメラ画像に対する目標領域の横のサイズの比率と縦のサイズの比率から、低い数値を選択する（カメラ画像 640x480、目標領域 320x120 のとき、 $640/320=2.0$ を選択する）。これは、ズームを行った後も目標領域全体をカメラ画像内に収めるためである。また、この時点における目標領域を抽出し、カメラ画像と同サイズに拡大した目標領域画像を得る。これは、第二段階において、ズーム後のカメラが取得する画像との位置ずれを求めするために使用する。目標領域とカメラ画像のアスペクト比が異なる場合は、目標領域画像の不足部分に黒帯を付加することで、目標領域画像とカメラ画像のアスペクト比を合わせる。ズームの実行と目標領域画像の取得の例を図 10 に示す。

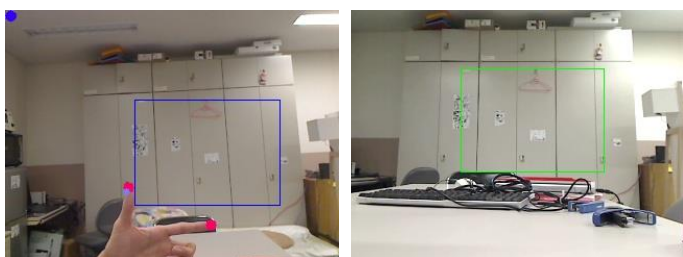


図 9. テンプレートマッチング

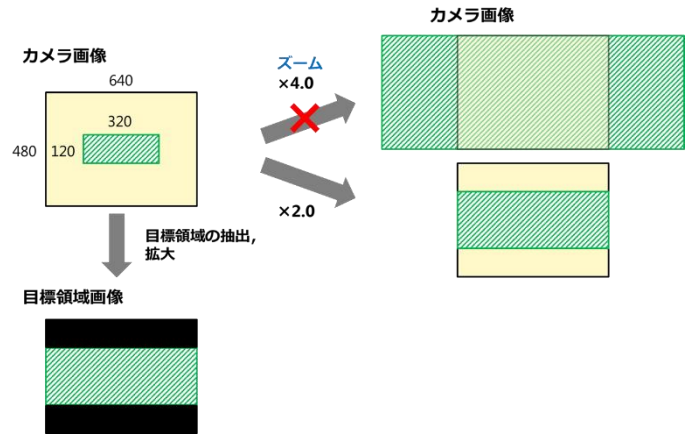


図 10. ズームの実行と目標領域画像の取得の例

ズームを実行した後、第二段階の「留まる」ための PTZ 制御を開始する。まず、カメラ画像と目標領域画像からそれぞれ特徴点を検出し、二枚の画像間で特徴点トラッキングを行う。トラッキングすることのできた特徴点から、二枚の画像間の位置ずれを求める。次に、位置ずれを解消するために、カメラをパン・チルトする。パン・チルトの値は、3.1 節で述べたカルマンフィルタによる推定手法から 1 度パン・チルトするときの特徴点の移動量を推定することで、位置ずれを解消するために必要な回転角度が求められる。ただし、肩載せカメラの位置・姿勢はユーザの身体の揺れに影響されて常に変動しており、期待通りの画像が取得出来るとは限らない。よって、パン・チルトした後は位置ずれを再度求め、制御を続行する。位置ずれが閾値以下であれば、撮影成功と判断し、カメラ画像を静止画保存する。閾値は、x 方向についてはパン、y 方向についてはチルトを 1 度行う毎の特徴点の移動量の半分と設定した。これは、パン・チルトした後に位置ずれが増大する状況を防ぐためである。

4. 評価実験

4.1. 実験手法

本システムにおける PTZ 制御手法の有効性を評価する実験を行った。被験者は頭部と右肩にカメラを装着する。右肩に装着するカメラは PTZ カメラである。被験者は直立した状態で頭部装着カメラの画像を確認し、図 11 に示した 5 つの領域の辺りを、それぞれ撮影領域に指定する。5 つの領域は順に、カメラ画像の左上、右上、左下、右下、中央を意味する。8 つの異なる環境における、頭部装着カメラ画像中から指定した撮影領域の画像と、肩載せ PTZ カメラから静止画保存した画像との類似度を求める。8 つの環境を環境 1 から環境 8 とし、図 12 に示す。類似度の計算は、撮影領域画像を静止画保存画像と同サイズに拡大し、両画像のゼロ平均正規化相互相関[6]の値を求める。両画像が完全に一致する場合の値は 1、相関が全くない場合の値は 0 である。値が正に大きいほど両画像の類似度が高いとわかる。

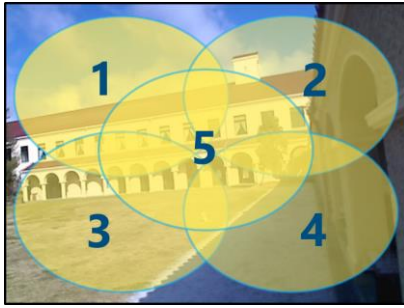


図 11. 指定する 5 つの領域



図 12. 環境 1～環境 8

本実験における頭部装着カメラの画像と肩載せカメラの画像の横サイズは 640pixel, 縦サイズは 480pixel である。撮影領域は、横サイズを 400 pixel, 縦サイズを 300 pixel に固定した。PTZ 制御の実行時において、被験者は常に右腕を下げている状態であり、故意に身体を動かすことはないものと想定する。

4.2. 結果

8つの環境で実験を行った結果を、それぞれ表 2 から表 9 に示す。表の横列の 1,2,3,4,5 は図 10 に示した頭部装着カメラの画像中の領域である。各領域の辺りを撮影領域画像に指定したときの、撮影領域画像と静止画保存画像の類似度と、撮影の成否を示している。撮影の成否は、ズームを行った後に何らかの原因で目標領域を見失う場合を失敗と判定する。また、肩載せカメラ画像を静止画保存したものの、誤った領域を収めていた場合も失敗と判定する。正しい領域の静止画保存を実行した場合を成功と判定する。

表 2 環境 1 での実験結果

	1	2	3	4	5
類似度	0.528	0.420	0.584	0.446	0.393
撮影の成否	成功	成功	成功	成功	成功

表 3 環境 2 での実験結果

	1	2	3	4	5
類似度	-	0.733	0.822	0.712	-
撮影の成否	失敗	成功	成功	成功	失敗

表 4 環境 3 での実験結果

	1	2	3	4	5
類似度	0.504	0.645	0.535	-	0.540
撮影の成否	成功	成功	成功	失敗	成功

表 5 環境 4 での実験結果

	1	2	3	4	5
類似度	0.496	0.681	0.809	-	0.762
撮影の成否	成功	成功	成功	失敗	成功

表 6 環境 5 での実験結果

	1	2	3	4	5
類似度	-	0.850	0.709	0.662	0.831
撮影の成否	失敗	成功	成功	失敗	成功

表 7 環境 6 での実験結果

	1	2	3	4	5
類似度	0.662	0.681	0.535	-	0.586
撮影の成否	成功	成功	成功	失敗	成功

表 8 環境 7 での実験結果

	1	2	3	4	5
類似度	0.642	-	0.735	0.368	0.574
撮影の成否	成功	失敗	成功	成功	成功

表 9 環境 8 での実験結果

	1	2	3	4	5
類似度	0.621	-	0.474	-	0.563
撮影の成否	成功	失敗	成功	失敗	成功

4.3. 考察

以下に考察を述べる。40 箇所中、計 30 箇所まで撮影に成功するという実験結果が得られた。環境 4 では右下の領域に対して撮影に失敗した。原因には、画像のハレーションが挙げられる。ハレーションとは画像内で特に強い光の当たった部分が白くぼやける現象のことである。強い光の当たった部分で特徴点トラッキングの誤りがあったために、パン・チルトが誤った回転角度で実行され、位置ずれが増大したと考えられる。位置ずれの大きさに比例して次のパン・チルトの回転角度と実行時間は増大するため、実行途中のフレームを誤って取得してしまう場合がある。図 12 に示すような実行途中のフレームからは特徴点がほとんど検出されないため、目標領域を見失い、撮影失敗と判定される。そこで、特徴点群から外れ値を持つものを除去す

れば、PTZ 制御の精度は向上すると考えられる。また、パン・チルトの回転角度に比例して、実行後にフレーム取得の待機時間を設ける必要がある。環境 5 の右下の領域に対する撮影では、誤った領域に対して静止画を保存していた。指定した撮影領域画像と静止画保存した画像を図 13 に示す。原因は、撮影領域と誤った領域とで画素値の分布が類似しており、テンプレートマッチングを誤ったためと考えられる。二つのカメラの位置関係から探索の範囲を限定することで、誤りを減らすことができると考えられる。また、撮影には成功したが、肩載せカメラの装着位置が肩の上からずれ落ちたために、撮影領域画像に対して静止画保存画像のロールが左にずれてしまい、画像の類似度が低くなる場合があった。例に、環境 4 の左上領域における撮影領域画像と静止画保存画像を図 14 に示す。ズーム倍率の調整と 2 次元アフィン変換を行うことで、画質をあまり低下させることなく、ロールのずれを補正した画像を抽出することができると考えられる。

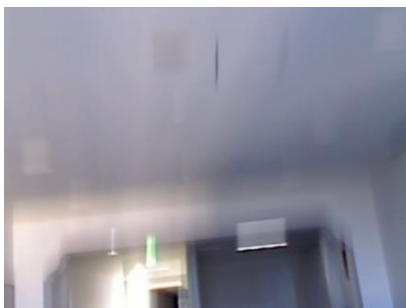


図 13. パン・チルト実行途中のフレーム



図 14. 環境 5 の領域 4 における撮影領域画像 (左) と静止画保存画像 (右)



図 15. 環境 4 の領域 5 における撮影領域画像 (左) と静止画保存画像 (右)

5. まとめと今後の展望

本研究では、ユーザがハンドフレームを用いて捉えた撮影領域画像と同等の構図を肩載せ PTZ カメラで撮影するシステムを作成した。まず、ユーザの頭部装着カメラの画像から特定のハンドジェスチャを検出することで、撮影領域画像の取得を行う。次に、肩載せカメラの画像中から撮影領域画像と同等の領域を探索し、発見した領域を目標領域と決定する。目標領域が肩載せカメラの全画面に表示されるようにパン・チルト・ズームを制御する。目標領域全体をカメラ画像内に収めたとき、カメラの画像を静止画保存する。

評価実験の結果、40 箇所中計 30 箇所 PTZ カメラの画像の静止画保存に成功し、本システムの有効性を示した。しかし、ハレーションの影響によって画像から特徴点を上手く検出することができず、撮影に失敗する場合があった。特徴点のトラッキングに誤りがあるとパン・チルトが誤った回転角度で実行されるため、位置ずれがより増大し、実行途中のフレームを取得してしまう。これを防ぐために、検出した特徴点群から外れ値をもつものを除去する手法を実装し、パン・チルトの回転角度に応じてフレーム取得の待機時間を設ける必要がある。また、画素値の分布が類似する、誤った領域を撮影してしまう場合があった。対応策に、ユーザの身体データ（腕の長さ等）を予め把握し利用することで、目標領域の探索範囲を幾何学的に限定することができないかと検討を進めている。これは、より小さな範囲を撮影領域に指定する場合にも、テンプレートマッチングの精度と処理時間を向上させるうえで有効であると考えられる。また、肩載せカメラはロール機能を持っていないため、撮影時に画像のロールのずれを解消することができない。カメラのズーム制御の再調整と画像の 2 次元アフィン変換を施すことで、ロールのずれを補正し、より類似度の高い画像を保存することができると考える。今後はこれらに対応し、撮影領域のサイズ変更やユーザの身体の揺れを加味した再検証を行いたい。

参考文献

- [1] 淵一馬, 高橋伸, 田中二郎. “ハンドジェスチャによる範囲選択手法を使った撮影システム”, 情報処理学会第 70 回全国大会, 2008.
- [2] 酒田信親, 蔵田武志, 興梠正克, 葛岡英明, マーク・ビリングハースト. “肩載せアクティブカメラ・レーザによる遠隔協調作業”, マルチメディア, 分散, 協調とモバイル(DICOMO2004)シンポジウム, pp.377-380, 2004.
- [3] Jamie Sherrah and Shaogang Gong. “Skin Colour Analysis”, http://homepages.inf.ed.ac.uk/rbf/CVonline/LOCAL_COPIES/GONG1/cvOnline-skinColourAnalysis.html, 2001. (2016/10/06 アクセス).
- [4] Elmar Mair, Gregory D. Hager, Darius Burschka, Michael Suppa, and Gerhard Hirzinger. “Adaptive and generic corner detection based on the accelerated segment test”, European Conference on Computer Vision (ECCV'10), 2010.
- [5] Gary Bradski and Adrian Kaehler (著), 松田 晃一 (訳). “詳解 OpenCV——コンピュータビジョンライブラリを使った画像処理・認識”, オライリー・ジャパン, 2009.