

超高速ミラー駆動パンチルトカメラを用いた 空中マルチタッチインターフェース

阿部 勇太^{1,a)} 入山 太嗣^{1,b)} 小室 孝^{1,c)} 島崎 航平^{2,d)} 石井 抱^{2,e)}

概要: 本論文では、超高速ミラー駆動パンチルトカメラを用いた高精度な手指骨格検出を利用し、端末から離れた位置からでもフリーハンドで、両手の5本指を用いたインタラクションを行える空中マルチタッチインターフェースを提案する。本インターフェースは、ユーザーの正面に操作ウィンドウを配置し、ウィンドウ内における手指の動きを利用して操作を行う。5本指のそれぞれを別の入力として扱うことで、片手のみで様々な操作を行うことや、指ごとに割り当てられた異なる操作を同時に行うことができる。手指の動きを利用したインタラクションには、それらの精細な動作情報を取得することが必要とされるため、超高速ミラー駆動パンチルトカメラを用いたカメラシステムを利用した高精度な両手指骨格検出を行う。本インターフェースを利用したアプリケーションの操作を通して、離れた位置からでもフリーハンドで両手の5本指を用いたマルチタッチによるインタラクションが可能になったことが確認できた。

1. イントロダクション

近年、セルフレジやキオスク端末などの利用者が自分で操作する端末の普及が進んでいる。これらの普及に伴い、不特定多数のユーザーが利用するシステムにおけるインターフェースについて考える必要がある。様々なインターフェースの中でも、コントローラーとなる装置を持たないフリーハンドによる操作インターフェースは、ユーザーの動きが直接操作へと反映されることから、操作が直感的でわかりやすく、初めて利用するユーザーでも簡単に操作方法を理解できるため、これらを操作するインターフェースとして用いられる機会が多い。特にタッチパネルによるタッチ操作や LeapMotion を用いた空中ハンドジェスチャー操作などは、フリーハンドによるインターフェースであることに加えて、デバイスを装着するなどのユーザーが操作に特別な準備を必要としないことから、身一つでだれでも簡単にシステムを利用することができる。しかしながらこれらのインターフェースは、操作を行うために端末に近づく必要があるという問題がある。

端末から離れた位置から利用可能なフリーハンドによるインターフェースとして、Kinect などの深度センサー [1]

や RGB カメラ画像 [2,3] によってユーザーの全身の骨格検出を行い、体の姿勢や動きを通してインタラクションを行うものがある。これらのインターフェースを利用することで、腕を振る動きに対応したカーソル移動や大まかな手の形を利用したハンドジェスチャー操作が可能であり、ポインティング [4] やオブジェクトの選択と操作 [5] を行うシステムが提案されている。しかしながら、腕の動きよりも精細な動作を利用してインタラクションを行う場合には、これらの動作情報を取得し、全身の骨格検出以上の高い空間分解能によるセンシングを行うことが必要となるため、マルチタッチのような5本指のそれぞれの動きを用いたインタラクションを行うことは難しい。

これらの関連研究を踏まえ、我々は超高速ミラー駆動パンチルトカメラと固定広視野カメラを用いた高精度な手指骨格検出を利用し、端末から離れた位置からでもフリーハンドで5本指を用いたインタラクションを行える空中マルチタッチインターフェースを提案する。本インターフェースでは、ユーザーの正面に操作ウィンドウを配置し、操作ウィンドウ内での手指の動きに応じて、ユーザーがその場から両手の5本指を用いたマルチタッチによって大型ディスプレイ全体の操作を行う。両手指の精細な動作情報を取得するために、高速カメラによるハイフレームレートでの撮影とミラーを用いた高速視線制御により、通常のパンチルトカメラの速度を超えて、高速に動き回る複数の対象を同時に追跡することが可能な、超高速ミラー駆動パンチルトカメラを利用する。固定広視野カメラで検出したユー

¹ 埼玉大学

² 広島大学

a) y.abe.796@ms.saitama-u.ac.jp

b) iriyama@mail.saitama-u.ac.jp

c) komuro@mail.saitama-u.ac.jp

d) simasaki@hiroshima-u.ac.jp

e) iishii@robotics.hiroshima-u.ac.jp

ザーの両手の位置を利用し、高速視線制御によって両手を高い空間分解能でセンシングすることで、両手指の精細な動作情報を取得し、高精度な両手指骨格検出を行う。

2. 空中マルチタッチインターフェース

2.1 概要

提案する空中マルチタッチインターフェースの概要図を図1に示す。本インターフェースは、ユーザーに情報を提示する大型ディスプレイと、ユーザーの動きを捉える超高速ミラー駆動パンチルトカメラと固定広視野カメラから成るカメラシステムによって構成される。カメラシステムがユーザーをセンシング範囲内に捉えると、そのユーザーを自動的に操作ユーザーとして登録し、操作ユーザーの正面に操作ウィンドウを配置する。ユーザーはこの操作ウィンドウを通じて、大型ディスプレイ上のアプリケーションにインタラクションを行うことができる。

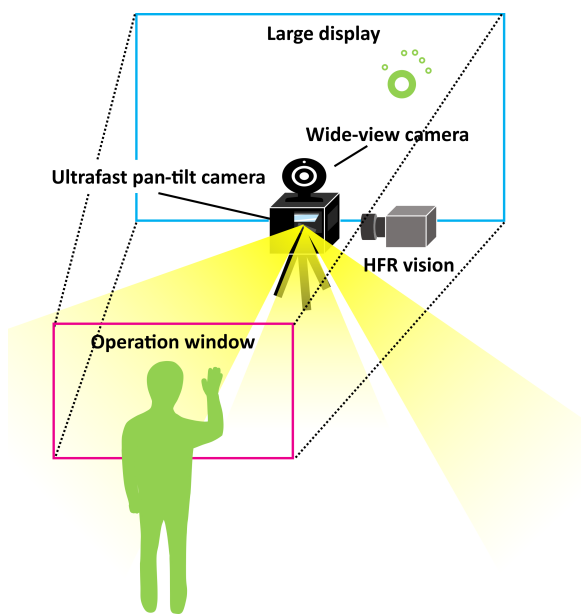


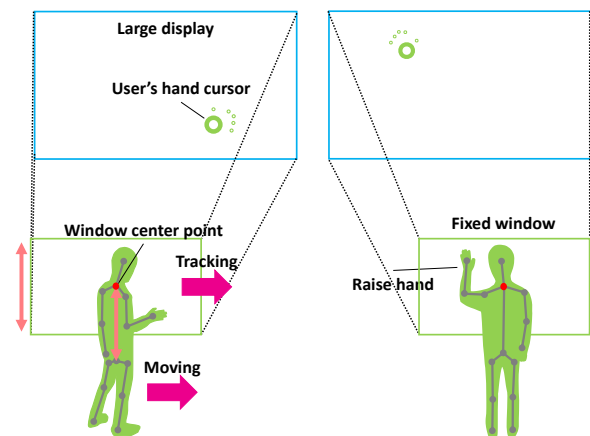
図 1: 空中マルチタッチインターフェース

操作ウィンドウ上の位置は、大型ディスプレイの画面上の位置と対応しており、ウィンドウ内でのユーザーの手指の動きに応じて、対応する場所にハンドカーソルを移動することができる。また本インターフェースでは、手を握った形を基本の姿勢として、指先を立てる動作によってタップ操作を行うことができる。タップ操作においては、インタラクションを行った指に応じて異なる操作を割り当てることが可能であり、これを利用して別々の操作を同時に行うことや、片手のみでも様々な操作を行うことが可能となる。

2.2 操作ウィンドウの配置と空間位置固定

空中マルチタッチインターフェースにおけるユーザーの

操作ウィンドウの配置について図2に示す。このウィンドウは、ユーザーの胸の位置を中心に配置され、通常時はユーザーの動きに合わせて追従を行う。ウィンドウの縦幅はユーザーの胸から腰までの距離を元に計算され、縦横比は大型ディスプレイの画面比と同様の比率で設定される。操作ウィンドウ全体が常にユーザーの手の届く範囲内に存在し続けることで、ユーザーはその場から動くことなく大型ディスプレイ全体を操作することが可能となる。また、本システムでは、端末を操作する意図を示すトリガーとして、ユーザーが手を挙げる行為を設定しており、トリガーとなる行動をとることによって操作ウィンドウがその場に固定される。ユーザーに追従して動き続ける操作ウィンドウが空間に固定されることで、追従時に比べて高精度での操作が可能となる。



(a) 移動するユーザーに合わ (b) 手を挙げたユーザーへの
せた操作ウィンドウの追跡 操作ウィンドウの固定

図 2: 操作ウィンドウの配置と空間固定

3. 超高速ミラー駆動パンチルトカメラを用いた高精度両手指骨格検出

空中マルチタッチインターフェースによる5本指を用いた操作を実現するためには、それぞれのユーザーの手指の精細な動作情報を取得し、高精度な手指骨格検出を行う必要がある。本研究では、超高速ミラー駆動パンチルトカメラを用いたカメラシステムを採用し、両手の位置の望遠映像を取得することで、ユーザーの両手指に対する空間分解能の高いセンシングを行う。

3.1 超高速ミラー駆動パンチルトカメラを用いたカメラシステム

超高速ミラー駆動パンチルトカメラを用いたカメラシステムの外観を図3に示す。

本カメラシステムは、操作対象となるユーザーの検出と

ユーザーの両手の位置の特定を行うための固定広視野カメラと、特定した手の位置へ注視を行い望遠映像を取得するための超高速ミラー駆動パンチルトカメラから構成される [6]。多対象注視用のミラー駆動パンチルトカメラにおけるカメラレンズには固定広視野カメラよりも焦点距離の長いレンズを使用する。ミラー駆動パンチルトカメラを、対象となる手ごとにマルチスレッドで高速視線制御することで、複数台の仮想望遠パンチルトカメラとして動作させる。



図 3: カメラシステムの外観

固定広視野カメラの映像を利用してユーザーの全身の骨格検出を行う。次に、固定広視野カメラの XY 画像座標系上の手の位置とミラーのパン角とチルト角を対応させる手法 [6] を利用することで、全身の骨格情報からユーザーの両手指の位置へミラー駆動パンチルトカメラを用いた注視を行い、手指の追跡・手指骨格検出を行う。

3.2 全身の骨格情報を用いたミラー駆動パンチルトカメラによるリアルタイム両手指骨格検出

提案手法では、図 4 のように、固定広視野カメラの広角視点を用いてユーザーの全身の骨格を検出し両手の位置を取得する。得られた手の位置情報を用いた、高速視線制御によって生成された複数台の仮想望遠パンチルトカメラの望遠視点により、ユーザーの手指へと同時に注視を行うことで、高い空間分解能を確保した両手指の情報取得を行い、その望遠視点の映像を利用して手指の骨格を検出する。全身と手指に対する骨格検出器には OpenPose [7] を採用する。

固定広視野カメラを利用して離れた位置に立つユーザーの全身骨格検出を行う際には、全身の骨格検出に必要な処理時間や検出精度が原因となり注視する位置にブレが発生する可能性がある。本研究では、手指への注視に使われたミラーのパンチルト角と固定広視野カメラの座標系上の位置情報を保持し、手指骨格検出結果と合わせることで、これらのブレを高精度な手指骨格検出の結果によって補正

し、正確な手指骨格の位置を取得する。

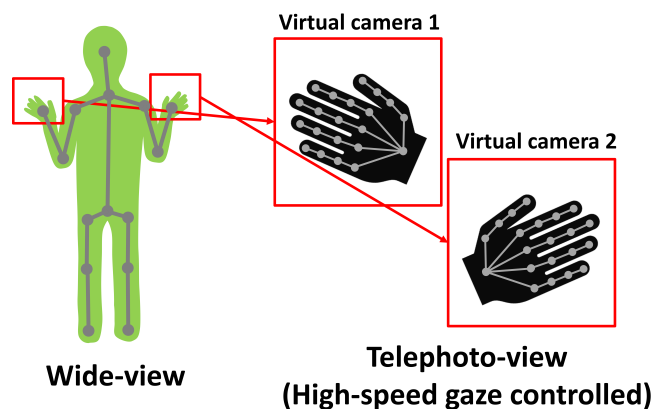


図 4: 全身の骨格情報を用いたミラー駆動パンチルトカメラによるリアルタイム両手指骨格検出

4. アプリケーション

本研究において提案した空中マルチタッチインターフェースを用いて、5 本指の精細な動きを利用したインタラクションが行えることを確認するため、マルチタッチによる操作を必要とする基本的なアプリケーションを実装し、操作を行った。図 5 に示すように、ユーザーはカメラシステムと映像が投影される大型ディスプレイから 5m 程度離れた位置から両手を利用してインタラクションを行った。

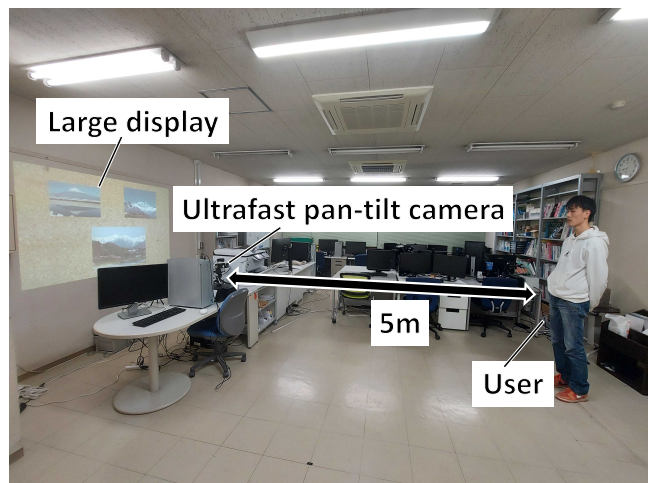
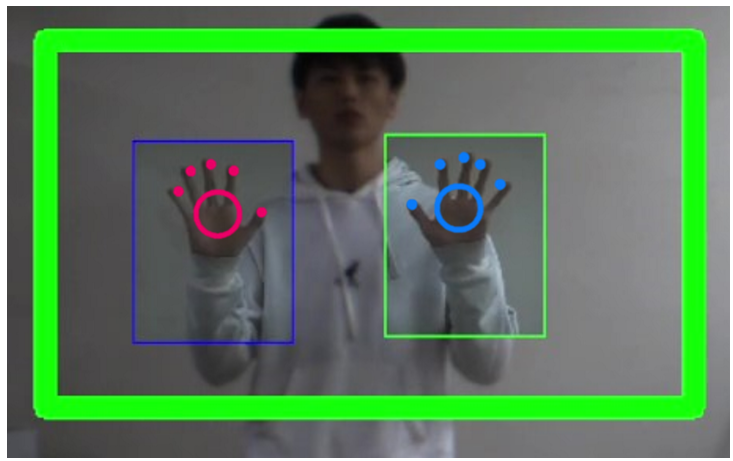


図 5: 操作環境

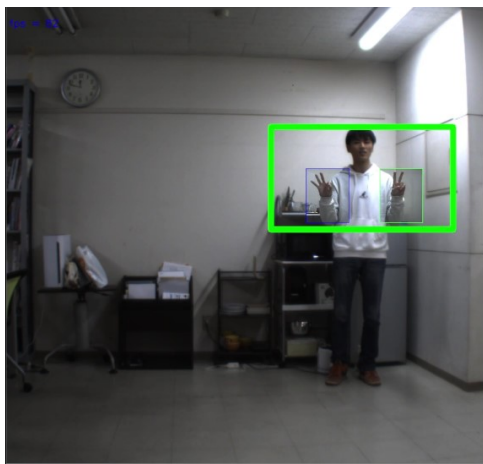
実験で使用したカメラについて、広視野カメラのレンズの焦点距離は 3.5 mm であり、その視野角は水平視野角 70.7°、垂直視野角 56.0° である。広視野カメラは 80fps として動作させた。また、超高速ミラー駆動パンチルトカメラモジュールにおけるレンズの焦点距離は 25 mm であり、その視野角は水平視野角 8.5°、垂直視野角 11.4° であ



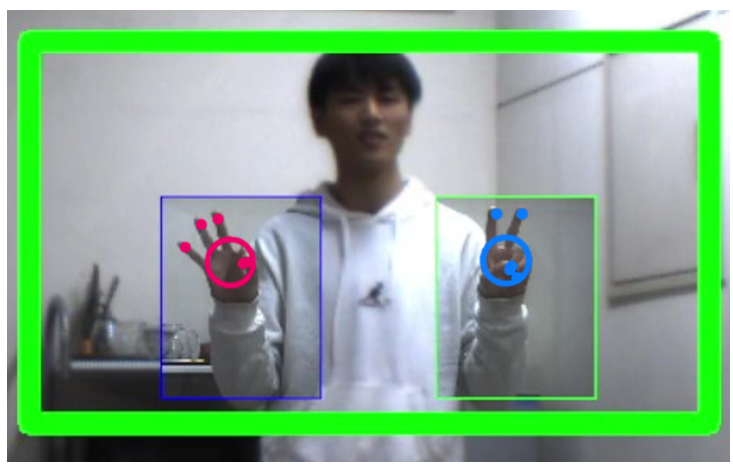
(a)



(b)



(c)



(d)

図 6: 操作ウィンドウを通したハンドカーソルの操作

る。ミラー駆動パンチルトカメラは 120fps で動作させており、これによって両手を取得するための二つの望遠視点はそれぞれ 60fps で動作するようになった。これらのカメラで前方 5m のシーンを観察した場合、広視野カメラの 1440×1080 画素の広角視点の映像は $5.6 \times 4.2\text{m}$ 、1 画素は 3.9mm である。また、超高速ミラー駆動パンチルトカメラモジュールの 480×640 画素の望遠視点の映像は $0.6 \times 0.8\text{m}$ 、1 画素は 1.25mm である。広角視点に対する望遠視点の解像度の倍率は 7.1 倍であった。超高速ミラー駆動パンチルトカメラモジュールのミラーの駆動範囲は、パンチルト角ともに 40° である。そのため、広角視点全体の映像から、このパンチルトカメラを利用して情報を取得できる範囲で映像を切り出したものを、今回は広角視点における映像として提示した。

図 6 では、それぞれのアプリケーションにおいて共通で利用される、操作ウィンドウを通じたハンドカーソルの操作について示す。ハンドカーソルは、両手の 5 本の指先を示す点と、手の中心部分を示す大きな円によって構成され

る。この時の手の中心位置には、中指の付け根となる部分の骨格情報を利用した。それぞれの指先に対応したカーソルは、インタラクションを行っている間、外側に同じ色の円が生成されることによってハイライトされる。図 6a や図 6c では、固定広視野カメラによって取得した映像上に操作ウィンドウを重畳したものを示す。常に操作ウィンドウがユーザーの正面となる位置に配置されていることが確認でき、これによってユーザーが位置を移動したとしても、同じようにハンドカーソルを操作することができた。図 6b や図 6d は、図 6a や 6c の映像から操作ウィンドウ周辺のみを切り出したものに、これらの操作を行っている際のアプリケーションにおけるハンドカーソルの動きを重畳したものである。操作ウィンドウ内でのユーザーの指先の位置とアプリケーションの空間上のハンドカーソルの位置はほとんど対応しており、高精度な手指骨格検出が行えていることが確認できた。また、手を開くような簡単な姿勢のみでなく、手を複雑な形にしていたとしても正しく検出を行い、その指先の位置へとカーソルが移動していることが分

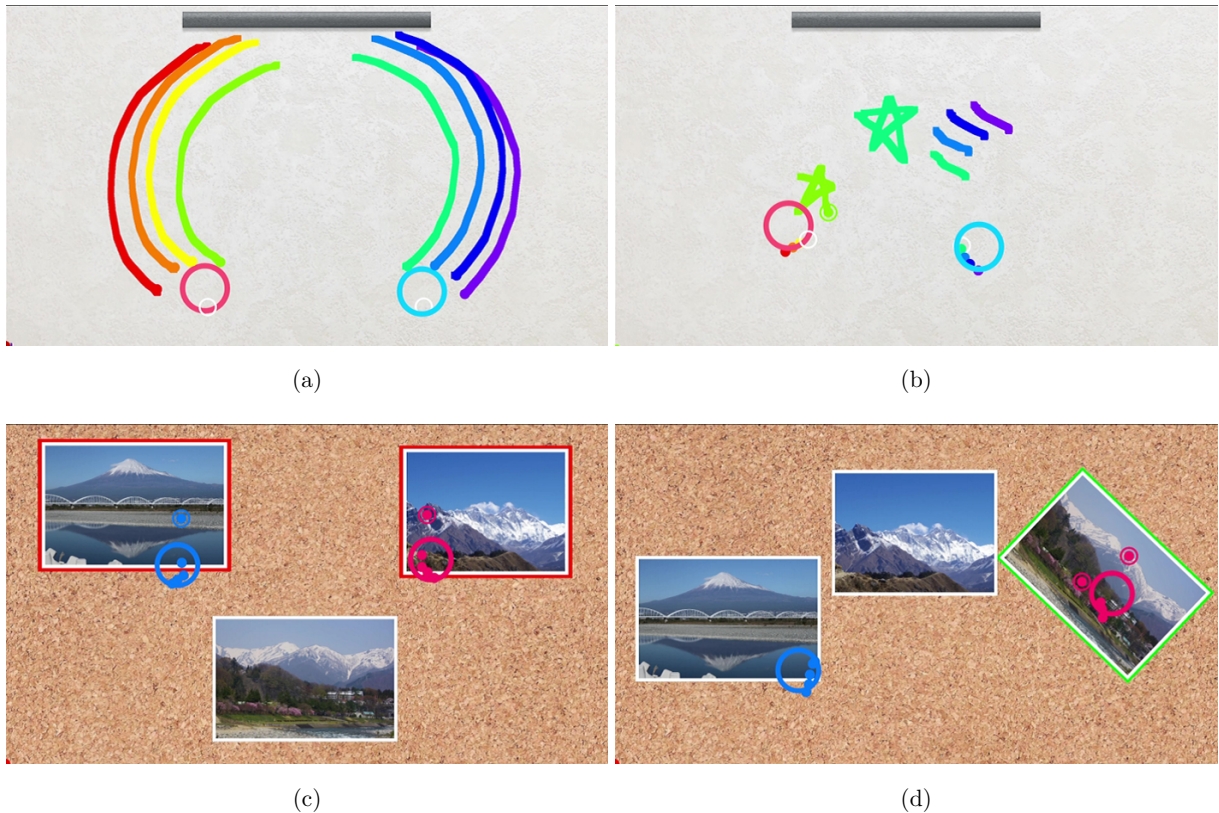


図 7: マルチタッチインターフェースを利用したお絵かきアプリと写真のレイアウトアプリの操作

かった。

マルチタッチ操作が必要な基本的なアプリケーションとして、お絵かきアプリと写真のレイアウトアプリの二つを実装した。お絵かきアプリでは、親指を除く両手のそれぞれの指に別々の色が割り当てられており、画面をタップすることで指に割り当てられた色を使って自由に線を描くことが可能である。また、親指には消しゴムが割り当てられており、描いた線に親指でタップすることで線を消すことができる。写真のレイアウトアプリでは、最初に数枚の写真が配置された状態から操作が開始し、写真の一点をタップすることで、タップした指先に追従して移動を行うことができる。また、二点をタップすることでタップした写真の拡大縮小、回転操作を行うことができる。

図 7 に、二つのアプリケーションのそれぞれを利用した際の画面の動きを示す。

図 7a や図 7b では、お絵かきアプリを利用して様々な操作を行った様子を示しており、複数の指を立てて異なる色の複数本の線を同時にひくことや、簡単な形の図形を描くことが可能であった。手を固定して指先のみを動かして図形を描いたり、指を握って一度線を書くのをやめて、別の場所で新しい線を引くという操作ができたことから、高精度な手指骨格検出により、立てられた指先の細かい動きを取得することや、どの指先を立てているのかを判別できていることが分かった。

図 7c では、写真のレイアウトアプリを操作して、二つの異なる写真のそれぞれに対して両手で一点タップを行い、同時に写真の移動を行った様子を示す。両手を使うことで、別々のオブジェクトに対して同時にインタラクションが可能であり、また離れた位置からでも動かずに画面内の全ての写真へインタラクションを行えることが確認できた。それぞれの指の精細な動きを捉えることで、図 7d のように、片手のみで写真をマルチタッチし拡大縮小や回転操作を行うなど、自由に写真のレイアウトを変更することができた。

5. まとめ

本論文では、超高速ミラー駆動パンチルトカメラと固定広視野カメラによって構成されるカメラシステムを用いた、空中マルチタッチインターフェースの提案を行った。本インターフェースは、ユーザーの正面に操作ウィンドウを配置し、ウィンドウ内での手指の動きに応じて、離れた位置から画面全体の操作を行うことを可能とする。またタップ操作において、片手のみで様々な操作を行うことや、指ごとに別の操作を割り当てて異なる操作を同時に行うことが可能である。超高速ミラー駆動パンチルトカメラを用いたカメラシステムを利用し、手指の精細な動作情報を取得することで、空中マルチタッチインターフェースにおける 5 本指のそれぞれを用いた操作に必要な高精度な手指骨

骨格検出を行った。本インターフェースを用いた基本的なアプリケーションの操作を通して、離れた位置からでも両手の5本指を利用したマルチタッチによるインタラクションが行えることが確認できた。

今後の課題として、複数人操作への対応や、その場合での操作ユーザーの数に応じた処理の最適化、オクルージョンが発生する姿勢の際に骨格検出の精度が低下するなどの情報抽出における問題の解決が考えられる。また本研究におけるインターフェースの操作についても、様々な操作手法と比較し、最適なものであるかやユーザーが自由に操作を行えるものであるかなどを定量的に評価することが必要である。

参考文献

- [1] Liu, J., Fujii, R., Tateyama, T., Iwamoto, Y. and Chen, Y.: Kinect-based gesture recognition for touchless visualization of medical images, *International Journal of Computer and Electrical Engineering*, Vol. 9, No. 2, pp. 421–429 (2017).
- [2] Nuzzi, C., Ghidini, S., Pagani, R., Pasinetti, S., Coffetti, G. and Sansoni, G.: Hands-Free: a robot augmented reality teleoperation system, *Proceedings of 2020 17th International Conference on Ubiquitous Robots (UR)*, pp. 617–624 (2020).
- [3] Liu, C. and Szirányi, T.: Real-time human detection and gesture recognition for on-board uav rescue, *Sensors*, Vol. 21, No. 6, p. 2180 (2021).
- [4] Tölgyessy, M., Dekan, M., Duchoň, F., Rodina, J., Hubinský, P., L'ubošChovanec: Foundations of visual linear human–robot interaction via pointing gesture navigation, *International Journal of Social Robotics*, Vol. 9, pp. 509–523 (2017).
- [5] Mäkelä, V., James, J., Keskinen, T., Hakulinen, J. and Turunen, M.: “It’s Natural to Grab and Pull”: Retrieving Content from Large Displays Using Mid-Air Gestures, *IEEE Pervasive Computing*, Vol. 16, No. 3, pp. 70–77 (2017).
- [6] Hu, S., Shimasaki, K., Jiang, M., Senoo, T. and Ishii, I.: A Simultaneous Multi-Object Zooming System Using an Ultrafast Pan-Tilt Camera, *IEEE Sensors Journal*, Vol. 21, No. 7, pp. 9436–9448 (2021).
- [7] Cao et al., Z.: OpenPose: Realtime Multi-Person 2D Pose Estimation Using Part Affinity Fields, *IEEE TPAMI*, Vol. 43, No. 1, pp. 172–186 (2021).