1. **Introduction**

- **Fingerspelling** is a tool to express a letter by a hand shape.
- **Main goal**: Detect and categorize fingerspelling in a continuous video as a mixture of fingerspelling sequences and irrelevant images.
- **Basic idea**: Divide a whole process into two-steps cascade process:
  1. **Spotting**: Segment and extract a fingerspelling sequence in an input video by utilizing temporal dynamic information.
  2. **Classification**: Classify the spotted sequence by utilizing 3D hand shape information.

**Dataset**:
- We recorded 15 fingerspelling classes by a depth camera.
- The hand region is extracted from the whole input image based on the depth map.
- We synthesized an input video, which continuously inputs fingerspelling and not fingerspelling sequences alternately.

**Evaluation index**:
- Spotting performance, Classification accuracy, and Recognition time.

2. **Classification process**

- **Step 1**: Spotsing process
  - Reference fingerspelling ($X$)
  - Segments a whole input video into fingerspelling sequences and irrelevant images.
  - Orthogonal Mutual Subspace Method (OMSM) with CNN features using hand shape information.

- **Step 2**: Classification process
  - Reference fingerspelling ($X$)
  - Spotted fingerspelling is classified by OMSM with CNN features using hand shape information.
  - The detailed procedure:
    1. 2.1. CNN features ($f_j^t$) and ($f_j^{t+1}$) are extracted from ($y_{t+1}$) and ($X$).
    2. Each class subspace ($S_j$) and an input subspace ($S_0$) are generated by applying PCA to the sets of CNN features.
    3. Orthogonal subspaces ($S_j$) and ($S_0$) are generated by applying orthogonal transformation to ($S_j$) and ($S_0$).
    4. The spotted fingerspelling is classified based on similarities between the input subspace ($S_0$) and reference subspaces ($S_j$).

**Conclusions**
- We proposed a fingerspelling recognition framework based on a complementary combination of TRCCA and OMSM with CNN features.
- We confirmed that our two-steps process significantly outperforms conventional one-step methods in terms of classification accuracy and recognition time.

**References**