論文 / 著書情報
Article / Book Information

論題(和文)	
Title(English)	User identification using Time-of-Flight camera image streams
著者(和文)	Gomez Caballero Felipe, 篠崎 隆宏, 古井 貞熙
Authors(English)	Felipe Gomez-Caballero, Takahiro Shinozaki, Sadaoki Furui
出典(和文)	情報処理学会創立50周年記念(第72回)全国大会, , No. 5X-8, pp. 2-615～2-626
Citation(English)	, , No. 5X-8, pp. 2-615～2-626
発行日 / Pub. date	2010, 3
権利情報 / Copyright	

5X-8

User identification using Time-of-Flight camera image streams

Felipe Gomez-Caballero　Takahiro Shinozaki　Sadaoki Furui

Graduate School of Information Science and Engineering, Tokyo Institute of Technology

1. Abstract

We propose a novel approach to identify users by comparing features extracted from image streams acquired from a Time-of-Flight camera. These features represent body landmarks which are detected and tracked over a small period of time in which the user is asked to perform specific movements. This information is later matched against a previously trained models corresponding to users movements. In order to measure the effectiveness of the proposed approach, we experimentally analyzed the performance of a prototype GMM-based identification system by using a small dataset.

2. Introduction

Capture and analysis of human motion is an active research area within the computer vision field due to its multiple applications and complexity. Common challenges are representation of real position/direction in a 3D space and foreground segmentation. This could be minimized by using a stereo image system or a Time-of-Flight (T-o-F) camera, being the latter system easier to setup, since it requires minimum calibration, but it also has lower resolution compared to standard RGB sensors. Although research on user identification based on monocular images exists, there are no research works that use the advantage of the depth images delivered by a T-o-F camera to perform this task [1][2].

In this paper we present a novel approach in which depth image streams obtained from a T-o-F camera are used to extract feature points in order to train probabilistic models, in a similar way to those used in speaker identification using speech signal, and use them to identify users by their movements.

3. Method

Our approach consists of three major phases, which are executed separately: feature extraction, model training and testing. The feature extraction and training phases are briefly described in this section while the testing phase is described in the next section.

3.1 Data acquisition and feature selection

The input to the feature extraction algorithm includes image streams acquired from a SwissRanger SR-4000 Time-of-Flight camera, captured at approximately 19.5 frames per second. The input is used to obtain three dimensional relative positions and direction vectors from five anatomical landmark points which are right/left elbow points, right/left shoulder points and a center point of the head. Moreover, center of mass of the body and an approximate area of the body are included in the feature set, resulting in a feature vector with 34 dimensions. These features were selected because our approach focuses in analyzing movements on the upper human body, specifically both arms motion.

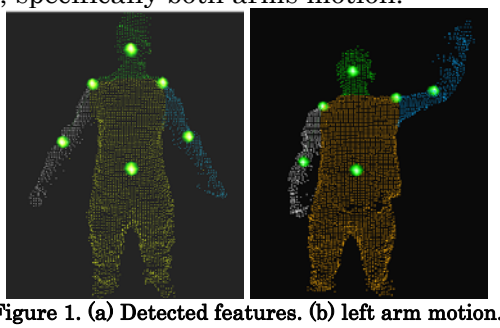

Figure 1. (a) Detected features. (b) left arm motion.

3.2 Feature extraction algorithm

Assuming that at the beginning of the process a person is facing the camera having an open arms position (*Fig.1* (*a*)), the algorithm first performs foreground – background segmentation by region growing [3] to separate the body from the rest of the scene, then it calculates the approximate center of mass of the body in order to find a rectangle defining the chest area which will be used as anchor for further segmentation of the limbs. Using this area as reference, the algorithm repeats region growing segmentation to find and separate regions corresponding to both arms and head, and then a low level image analysis, taking depth image property into consideration, is performed to obtain each feature point. This process is repeated for each frame of the image stream, using previously found feature points as seeds on each step of the region growing segmentation. The algorithm also uses a

simple Kalman filter to track the feature points and use position estimation in case of ambiguities which can result in fewer detected points.

3.3 Model training

A GMM model of each user is trained by using the EM parameter estimation method with a Cross Validation Gaussian optimization method, which automatically controls the number of Gaussian components in a GMM [4].

4. Experiments

Experiments were performed using a small dataset, consisting of 13 users with 20 samples each, including image streams recorded from a single Time-of-Flight camera depicting arm movements. For this test the sequences depicting single arm raising movement were used in the sense of trying to simulate a scenario in which an automatic system asks the user to raise his/her arm. Left arm movement (*Fig.1* (*b*)) was chosen because we consider it more effective, with the fact that most of the users included in the dataset are right handed and the left arm movements are difficult to be imitated by other users.

Four different feature subset setups were used in order to test the performance and significance of the features. Each setup was made as follows (number in parenthesis represents the feature vector dimension): {A} left shoulder relative point and direction vector (6), {B} left elbow relative point and direction vector (6), {C} left shoulder and elbow relative points and direction vectors (12), {D} shoulder and elbow direction vectors from both arms (12).

Also, the full feature set for the first dataset and data recorded in a different session were used in order to test the robustness of the features analyzed in this experiment. The second session data were recorded approximately two weeks after the first recording and include 11 out of the 13 users, with 5 samples each (2nd S.).

5. Results and discussion

Leave-one-out cross validation (LoOCV) with 20 folds was used to evaluate recognition performance. Average likelihood per frame was used to decide if the user was correctly identified. Graph in Fig. 2 shows the average accuracy of each subset over the 20 folds of the LoOCV.

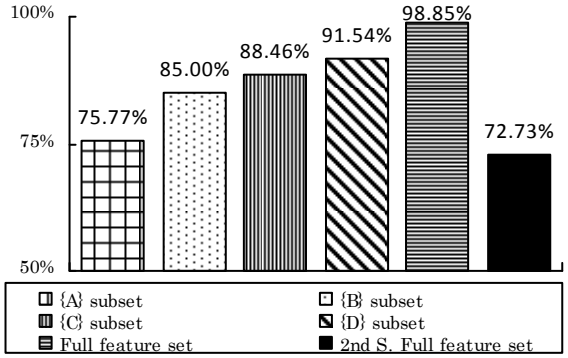

Figure 2. Recognition accuracy for each setup.

It is shown that by using the dataset which includes only data from the same recording session, performance increases if we include more features, hence we can rely in the distinctiveness of the chosen features. However, we also find a decrease in accuracy when using data recorded in a different session. This was mainly due to errors in feature point detection for 3 users, thus creating ambiguities that our feature extraction algorithm cannot handle, hence affecting the distinctiveness of the extracted features. Although this problem caused 0% accuracy for the 3 users, the accuracy for rest of the users was mostly 100%.

6. Conclusions and future work

We have shown a novel and promising approach for user identification and demonstrated its effectiveness by an experimental setup. We found some cases in which our proposed feature extraction algorithm failed to process ambiguities in the input data, so we would like to improve the performance of the algorithm to cope with this problem. Also, we speculate that by including both hand points and dynamic features, and giving extra weight to distinctive features, we can increase the accuracy of the proposed approach. Moreover, we would like to experiment with a larger data set and recordings with bigger time lapses.

7. References

[1] A. Kolb, E. Barth, R. Koch. "*ToF-sensors: New dimensions for realism and interactivity*". CVPRW '08, pp 1-6 (2008).
[2] T.B. Moeslund et al. "*A survey of advances in vision-based human motion capture and analysis*". Computer vision and image understanding, pp 90-126, (2006)
[3] L. Bianchi et al. "*Tracking without Background Model for Time-of-Flight Cameras*". LNCS; Vol. 5414, pp 726-737 (2009)
[4] T. Shinozaki, S. Furui, and T. Kawahara. "*Aggregated cross-validation and its efficient application to Gaussian mixture optimization.*" Interspeech 2008, pp.2382-2385 (2008).