

摘要

手语识别研究致力于通过计算机将手语翻译成文字或者语音,以方便聋人与健听人之间的交流和帮助聋人更好地融入社会。

依手语数据获取方式的不同,手语识别研究通常分为基于数据手套的手语识别和基于视觉的手语识别两个类别。二者当中,基于视觉的手语识别研究,由于应用起来更为方便和自然,备受研究者所关注。然而,大多数基于视觉的手语识别研究限定摄像机的捕获视角,通常限定为正面视角。限定捕获视角意味着手语者只能在特定的空间以特定的朝向执行手语,这严重地限制着手语者的自由。

本文针对单摄像机应用中、一定范围内视角无关的手语识别,也即观测手语样本的捕获视角在一定范围内任意且未知情况下的手语识别进行研究,以在一定程度上解除对摄像机捕获视角的限定,从而方便用户的使用。这里的一定范围,是指摄像机的光轴方向与手语者正面朝向之间的角度限制在 45° 范围之内。数据规模初步定于中等词汇集。

重点研究如下两方面内容:

一是基于视角相关特征的视角无关手语识别研究。基于虚拟立体视约束,即“同一手语不同捕获视角下的两个样本,对齐之后可解释为由某一虚拟的立体视觉系统同步捕获的一个样本对”,本文提出一种新颖的、使用视角相关特征的视角无关手语识别框架。此框架将两个手语序列的匹配问题转化为极几何中的一个验证问题,通过验证“两个手语序列是否能被解释为由某一虚拟立体视觉系统同步捕获的一个样本对”来完成识别。由于是直接基于特征点的图像坐标进行识别,此框架提供了一种使用视角相关特征而实现视角无关手语识别的可能性。基于此框架,本文提出了三种具体的使用视角相关特征的视角无关手语识别方法,分别是基于基础矩阵唯一性的视角无关手语识别方法、基于证据理论的验证基础矩阵唯一性的视角无关手语识别方法以及基于复合帧对应的视角无关手语识别方法。实验显示了这些方法的有效性。

二是短时数据缺失下稳定而有效的视角无关手语识别方法研究。所谓数据缺失,是指手语样本中某些时刻只能提供很少的有用特征以供匹配之用或者不能提供任何有用特征的情况。很多因素都能导致数据缺失的出现,比如自遮挡、成像因素导致的图像模糊、以及特征提取算法的不完善等因素。除此之外,在视角无关的手语识别中,同一手语不同捕获视角的样本之间可视特征的不尽一致,也能够导致数据缺失的出现。数据缺失影响识别算法的有效性和可行性。本文针对短时数据缺失下稳定而有效的视角无关手语识别方法进行重点研究,提出了基于运动元顺序出现单应性的视角无关手语识别方法以及基于基础矩阵的局部采样(Sample)加全局验证(Consensus)的视角无关手语识别方法。这两种方法基于多帧图像来进行序列匹配的考察,由于多帧图像可以提供更多的有用特征,这两种方法能够有效地处理短时数据缺失情况。实验显示了这两种方法的有效性。

值得指出的是,本文提出的基于虚拟立体视约束的视角无关识别框架和因之而提出的各种视角无关识别方法不仅仅适用于手语识别,还能应用到更为广泛的领域,比如视角无关的动作识别和刚体运动分析等。

Abstract

Sign language recognition aims to translate sign language into text or speech by computer, so as to facilitate the communication between the deaf and the hearing people and help the deaf or hard-of-hearing better integrate into the society.

According to data collection of sign language, sign language recognition are generally divided into two major categories: dataglove-based sign language recognition and vision-based sign language recognition. Since the vision-based method is more convenient to the end-users than dataglove-based method, it attracts more attention of the researchers. However, most of the current methods require a specific view of the signers, generally the frontal view. This constraint means that the signers can only perform with specific location and orientation, and limits the freedom of the signer.

The thesis aims to achieve viewpoint free sign language recognition within a certain scope with only one camera, so as to remove the restriction of a specified view and provide convenience for the user. The scope now covers a range from 0° to 45° from the point of view of the angle between the optical axis of the camera and the orientation of the signer. The size of sign language vocabulary is preliminarily set to medium.

The thesis focuses on the following two aspects:

(A) Viewpoint free sign language recognition based on viewpoint dependent features. Based on the FSV (Fictitious Stereo Vision) constraint that all corresponding frame pairs between two sequences of the same sign but from different viewpoints can be explained as captured synchronously in some fictitious stereo vision system, the thesis proposes a novel FSV recognition framework for viewpoint free sign language recognition. The proposed FSV recognition framework converts the recognition task to a verification task in the framework of epipolar geometry and achieves recognition by verifying whether two sign sequences can be explained as captured synchronously in some fictitious stereo vision system. Because the FSV recognition framework employs the image coordinate of feature points, it is a way to viewpoint free sign language recognition from viewpoint dependent features. Based on the proposed FSV recognition framework, the thesis proposes three methods for viewpoint free sign language recognition, including the method based on the uniqueness of fundamental matrices, the method of employing evidence theory to verify the uniqueness of fundamental matrices and the method based on the correspondence between the compounded frame pair. Experiments show the efficiency of the proposed three methods.

(B) Viewpoint free sign language recognition under short-duration data deficiency. Data deficiency refers to the case that some frames in a sign language sequence can provide only small number of efficient features or no efficient feature for matching. Data deficiency may be caused by many factors such as self-occlusion, the image blur and the imperfection of the feature extraction algorithm. Besides, in the application of viewpoint free sign language recognition, the observable feature set of an observation sample may be different from that of its matched template sample, which may also cause data deficiency. Data deficiency may affect the efficiency and feasibility of recognition algorithm. The thesis emphasizes the case of short-duration data deficiency and aims to achieve robust and efficient methods for viewpoint free sign language recognition under such cases. In the light of more frames providing more features, the thesis proposes two novel methods for short-duration data deficiency, including the method based on the

homography of tiny motions and the fundamental-matrix-centered Sample-Consensus method. Experiments show the efficiency of the proposed two methods.

It is worth noting that the proposed FSV recognition framework and all the proposed methods for viewpoint free sign language recognition not only fit for sign language recognition, but also can be applied to more broad fields such as viewpoint free motion recognition and rigid-motion analysis.