

摘要

随着计算机在现代社会影响的迅速扩大，传统的基于鼠标和键盘的人机交互技术越来越显示出它们的局限性，所以研究多模式人机接口技术在现实生活中变得越来越重要。手语识别作为多模式人机接口领域的一项重要组成部分，已经吸引了越来越多的专家和学者们的注意。手语识别的目标就是通过计算机提供一种有效、准确的机制将手语翻译成文本或语音，使得聋人和听力正常人之间的交流变得更方便、快捷。

手语识别目前所面临的主要挑战是如何建立合适的模型来解决非特定人、大词汇量、连续手语，以及怎样利用语言模型来提高手语识别率这四个问题。一个系统能够非常准确地识别非特定人、大词汇量、连续手语，而没有引入人工的停顿，对于方便、自然的人机接口具有深远的影响。因此，对上述四个问题的研究使得手语识别系统具有更强的鲁棒性和友好性，从而推动识别系统得到更广泛的应用。

本文针对手语识别中的四个问题：非特定人、大词汇量、连续手语、如何利用语言模型来提高手语识别率，分别建立合适的模型来解决这些问题：

1. 针对非特定人手语识别的特点：1) 数据多且差异大，导致模型训练难收敛；2) 从不同人数据中提取出有效的共同特征缺乏，本文提出了自组织特征映射/隐马尔可夫模型 (SOFM/HMM) 相结合的模型。该模型以 SOFM 隐式地提取不同人特征作为连续 HMM 的输入，将数据变换成一个紧凑、重要的低维表示形式，该形式能够更好地被 HMM 的发射概率模型化。它们的模型参数是在统一的全局优化准则下训练得到的。实验结果表明，该模型比传统的 HMM 模型识别率提高近 3—5%，较好地解决了非特定人识别问题。

2. 为了克服大词汇量识别所带来的时间复杂性增加的困难，在 SOFM/HMM 模型的基础上，提出具有异构分类器的模糊决策树用于大词汇量的手语识别。由于不同的特征对于手势词具有不同的模式区分性，因此，本文提出了异构的分类器来分层决策手语的属性。基于高斯混合模型的单双手分类器和基于有穷状态机的手形分类器首先被用来消除不可能的候选，然后在底层仅包含很少一部分候选词集的非叶子节点上，使用 SOFM/HMM 方法进行分类。实验结果表明，该方法在大词汇量非特定人手语识别中比单独使用 SOFM/HMM 方法大大减少了识别的时间，大约 11 倍，同时也相应地提高识别率 0.95%。

3. 在连续手语识别中，面临的主要挑战是如何减轻相邻手语词之间运动插

入的影响。本文从基于分割和建立过渡模型的思想入手，分别提出基于精简循环网/隐马尔可夫模型（SRN/HMM）相结合的模型和基于过渡模型的方法进行连续手语识别。1) 基于 SRN/HMM 模型的方法是将连续手语识别问题分解成各孤立词识别的分治方法。把改进的 SRN 作为连续手语的段边界检测器，SRN 的分段结果作为 HMM 框架中的状态输入，在 HMM 框架里使用网格 Viterbi 算法搜索出一条最佳的手语词路径。2) 基于过渡模型的方法是将词与词之间的过渡动作也建立相应的模型来进行识别。为了克服词与词之间大量的过渡模型，本文提出了时序聚类算法，它能将相似的过渡动作聚成一类，从而增强过渡模型的推广性，同时避免训练数据的稀疏问题。实验结果表明，基于过渡模型的方法在大词汇量连续手语识别中取得了较好的效果。

4. 在统计语言模型中，如何将多种语言学知识融入到一个统一的框架下，作为长距离的约束关系来提高手语识别率是一个挑战。本文提出了一个融入语言学结构知识的改进最大熵语言模型。该模型把基本短语的结构知识与 Trigram 结合，Trigram 作为词之间短距离的约束，而用分析出基本短语的结构知识来表示句法结构中长距离的约束关系；将语法、语义、词汇这些语言学知识统一在最大熵的框架下。实验结果表明，该模型比 Trigram 在分支度上提高 24% 左右。同时提出了手语同义词的概念，通过手语同义词的扩展，将改进的最大熵语言模型作为连续手语识别的后处理，有效地提高了手语识别的性能。

关键词 手语识别；手语模型；语言模型；人工神经网络；隐马尔可夫模型

Abstract

With the widespread use of the computer in modern society, traditional human machine interaction (HCI) technologies based on mouse and keyboard show their increasing limitations. Thus, the research on multimodal HCI becomes more and more important in real life. Sign language recognition (SLR), as one of the important research areas of HCI, has spawned more and more interest in HCI society. The goal of sign language recognition is to provide an efficient and accurate mechanism to transcribe sign language into text or speech so that communication between deaf and hearing society can be more convenient.

The major challenges that SLR faces now are how to build the effective models for signer-independent, large vocabulary, continuous sign and how to improve the SLR accuracy through language models. The ability to accurately recognize signer-independent, large vocabulary and continuous sign language for a system, without the introduction of artificial pause, has a profound influence on the naturalness of the human-computer interface. Therefore, the research on these four challenges can promote the development of more robust and convenient system, so that the SLR system will be widely used in the various situations.

Aiming at four issues: signer-independent, large-vocabulary, continuous, and how to improve SLR performance using language models, the dissertation proposes the corresponding models to solve them:

First, aiming at signer-independent SLR characteristics: 1) the model convergence difficulty caused by noticeable distinctions among different people signs, 2) the lack of effective features extracted from different signers' data, self-organizing feature maps/hidden Markov models (SOFM/HMM) is proposed. This model uses SOFM as an implicit feature extractor for continuous HMM and their parameters are trained simultaneously in a global optimization criterion. SOFM transforms input signs into significant and low-dimensional representations that can be well modeled by the emission probabilities of HMM. Experimental results show that the proposed model increases the recognition accuracy about 3-5% compared with conventional HMM, and well solve signer-independent SLR issue.

Second, to overcome the difficulty of the huge time complexity due to a variety

of recognized classes, a fuzzy decision tree with heterogeneous classifiers based on SOFM/HMM is proposed for large vocabulary SLR. As each sign feature has the different discrimination to gestures, the corresponding classifiers are presented for the hierarchical decision to sign language attributes. One- or two- handed classifier base on Gaussian mixture models and hand shape classifier based on finite state machine are first used to progressively eliminate many impossible candidates, and then SOFM/HMM classifier is proposed as a special component of fuzzy decision tree to get the final results at the last non-leaf nodes that only include few candidates. Experimental results on a large vocabulary demonstrate that the proposed method dramatically reduces the recognition time by 11 times, and also improves the recognition rate about 0.95% over single SOFM/HMM.

Third, the major challenge for continuous SLR is how to alleviate the effect of movement epenthesis between two adjacent signs. Based on two strategies: segmentation-based and modeling extra movements, this dissertation respectively proposes simple recurrent networks/hidden Markov models (SRN/HMM) and transition model for continuous SLR. 1) SRN/HMM is a divide-and-conquer recognition method that breaks down the problem of continuous SLR into individual sign recognition. This method applies the improved SRN to segment continuous sign language, and the outputs of SRN are taken as the HMM states in which the lattice Viterbi algorithm is employed to search the best word sequence. 2) Transition model is to build similar sign models for transition movements between adjacent signs in SLR. To overcome a large amount of transition movements, the temporal clustering algorithm is proposed to cluster them. The clustered models can improve the generalization of transition movement models, and are very suitable for large vocabulary continuous SLR. Experiments show that large vocabulary continuous SLR based on transition movement models has good performance.

Fourth, in statistical language models, how to integrate diverse linguistic knowledge into a general framework as long-distance dependencies to improve SLR performance is a challenging issue. In this dissertation, an improved language model incorporating linguistic structure into maximum entropy framework is presented. The proposed model combines trigram with structure knowledge of base phrase in which trigram is used to capture the local relation between words, while structure knowledge of base phrase is considered to represent the long-distance relations

between syntactical structures. The knowledge of syntax, semantics and word is integrated into the maximum entropy framework. Experimental results show that the proposed model improves by 24% language model perplexity over trigram. Furthermore, the concept of sign language synonymy is proposed. After the enlargement of sign language synonymy, the proposed model is integrated as the post-process of continuous SLR, and effectively improve sign language recognition rate.

Keywords sign language recognition; sign model; language model; artificial neural network; hidden Markov model