

Desenvolvimento de um Sistema de Reconhecimento Automático de Língua Brasileira de Sinais

Beatriz Teodoro¹, Luciano A. Digiampietri¹

¹Escola de Artes, Ciências e Humanidades – Universidade de São Paulo (USP)
Av. Arlindo Bértio, 1000 – Ermelino Matarazzo – 03828-000 – São Paulo – SP – Brasil

{beatriz.teodoro, digiampietri}@usp.br

Abstract. *The sign language recognition is an important research area that aims to increase the integration of deaf people into society, but that is in the early stage of development. The objective of this project is to develop an information system for the automatic recognition of Brazilian Sign Language (BSL), in order to simplify the communication between deaf people talking in BSL with who have no knowledge of this sign language.*

Resumo. *O reconhecimento de língua de sinais é uma importante área de pesquisa que tem como objetivo aumentar a integração das pessoas surdas e/ou com deficiência auditiva na sociedade, mas que está em estado inicial de desenvolvimento. O objetivo deste projeto é desenvolver um sistema de informação para o reconhecimento automático de Língua Brasileira de Sinais (LIBRAS), de forma a simplificar a comunicação entre surdos conversando em LIBRAS e ouvintes que não conheçam esta língua de sinais.*

1. Introdução

Nos últimos anos, é possível observar que o empenho em facilitar a comunicação entre surdos e pessoas que não conhecem uma língua gestual tem aumentado, mas ainda há poucos ambientes acessíveis para os surdos. A inclusão dos deficientes auditivos na sociedade (majoritariamente ouvinte) tem enfrentado a falta de conhecimento dos ouvintes sobre línguas de sinais, que são línguas gestuais usadas por surdos e deficientes auditivos para se comunicarem. Esta falta de conhecimento torna extremamente difícil a comunicação entre surdos e ouvintes. Além disso, o reconhecimento e tradução de línguas de sinais são áreas bastante complexas e que estão em estado inicial de desenvolvimento.

Nas línguas de sinais, tanto na brasileira como em outras [Albres 2010, Quadros and Karnopp 2004], existem sinais estáticos para indicar algumas palavras e letras, porém muitos dos sinais são dinâmicos (envolvem não só a configuração das mãos, mas também o movimento). Desta forma, para o processamento automático de uma conversa em língua de sinais é necessário o processamento de um vídeo (conjunto sequencial de imagens) e não somente das imagens individualmente.

Este projeto visa a estender trabalhos prévios [Digiampietri et al. 2012, Teodoro 2012] sobre reconhecimento de Língua Brasileira de Sinais (LIBRAS), através da implementação de uma nova ferramenta, que tem como objetivo reconhecer e traduzir sinais dinâmicos, com a finalidade de simplificar a comunicação entre surdos conversando em LIBRAS e ouvintes que não conheçam esta língua. O reconhecimento será realizado através do processamento de imagens e vídeos digitais de pessoas se comunicando em

LIBRAS, sem o uso de luvas coloridas e/ou luvas de dados e sensores ou a exigência de gravações de alta qualidade em estúdios, focando em sinais que utilizam apenas as mãos.

2. Conceitos Básicos e Trabalhos Correlatos

As línguas de sinais são línguas naturais utilizadas como forma de comunicação entre deficientes auditivos. Cada país possui a sua própria língua de sinais que, assim como as demais línguas, é influenciada pela cultura nacional. Existem cinco parâmetros relacionados à realização de sinais em língua sinalizada: (a) configuração da mão (neste trabalho estão sendo consideradas 63 configurações diferentes¹); (b) movimento; (c) ponto de articulação (local do espaço de sinalização onde o sinal é realizado) [Stokoe 2005]; (d) orientação da palma da mão [Klima and Bellugi 1979]; e (e) expressões não manuais (por exemplo, faciais). Neste projeto, pretende-se utilizar os parâmetros (a), (b), (c) e (d) para o reconhecimento de sinais.

Além dos parâmetros que formam os sinais, vale ressaltar também os métodos de aquisição de sinais. Atualmente, existem duas abordagens principais na área de reconhecimento de língua de sinais: a abordagem visual, onde os dados são obtidos através de uma câmera de vídeo, e a abordagem baseada em dispositivos eletromecânicos, como luvas de dados e sensores. A primeira abordagem é extremamente adequada para ser aplicada no dia-a-dia, sendo mais conveniente para o usuário, mas exige um pré-processamento mais sofisticado. Na segunda abordagem existem desconfortos e limitações para o usuário utilizar uma luva de dados e/ou sensores, porém esta abordagem facilita o reconhecimento, evitando problemas enfrentados na primeira abordagem, como o de segmentação e rastreamento das mãos. No entanto, a fim de diminuir a complexidade do pré-processamento das imagens obtidas através da abordagem visual, muitos trabalhos utilizam ambientes controlados e/ou luvas coloridas na gravação dos vídeos [Akmeliawati et al. 2007], que facilitam consideravelmente o reconhecimento dos sinais. Neste trabalho, será utilizada a abordagem visual, mas sem a utilização de luvas coloridas. Desta forma, o reconhecimento de LIBRAS será feito sobre vídeos gravados de maneira pouco intrusiva e sem serem gravados especificamente para facilitar o reconhecimento automático.

Diversos trabalhos relacionados ao reconhecimento de línguas de sinais foram publicados ao longo dos últimos anos. A fim de vislumbrar o estado da arte sobre reconhecimento de línguas de sinais, foi realizado um processo de revisão sistemática sobre o assunto, focando nos trabalhos que utilizam a abordagem visual, mas por questões de espaço, no restante desta seção será apresentado apenas um pequeno resumo da revisão, ressaltando as principais características encontradas.

Nota-se a partir dos trabalhos analisados que o reconhecimento de sinais dinâmicos em língua de sinais ainda é pouco desenvolvido. A maioria dos trabalhos encontrados apresentam resultados bastante satisfatórios ao reconhecer a configuração de mão (ou letra ou número) que está sendo sinalizada em uma única imagem [Dias et al. 2004], porém não foram encontrados resultados tão bons nos trabalhos que consideram sinais dinâmicos (que são aqueles compostos por sequências de configurações

¹Não há um consenso sobre quantas e quais configurações de mão são mais adequadas para a descrição das línguas de sinais. Há trabalhos sobre LIBRAS que utilizam de 46 a 73 configurações [Ferreira et al. 2011]. Em nosso trabalho optamos por utilizar as 63 configurações presentes no Dicionário Digital do Instituto Nacional de Educação de Surdos (<http://www.ines.gov.br/>)

de mão e movimento).

Outro ponto que chamou a atenção na revisão é a pequena presença de sistemas completos, que utilizam todos os cinco parâmetros relacionados à realização de sinais, a maioria dos trabalhos selecionados focam apenas em dois ou três parâmetros. Também notam-se poucos trabalhos que estudam expressões não manuais em reconhecimento de sinais. Este parâmetro em particular é muito importante ao se pensar em reconhecimento de sinais e, até o momento, foi pouco trabalhado. Por fim, outro ponto que também chama a atenção, é a ausência de trabalhos que possam ser generalizados para o reconhecimento de sinais em diversas línguas de sinais. A grande maioria das técnicas utilizadas nos trabalhos encontrados é focada em línguas de sinais específicas, sendo estes pontos, possíveis lacunas a serem exploradas em trabalhos futuros.

3. Caracterização da Contribuição

A principal contribuição deste trabalho é o sistema de informação para o reconhecimento automático de LIBRAS proposto, capaz de analisar vídeos de pessoas se comunicando em LIBRAS, reconhecer os sinais expressos nesses vídeos e apresentar as palavras correspondentes em língua portuguesa. Pretende-se facilitar, desta forma, a comunicação entre surdos e/ou deficientes auditivos e pessoas que não conhecem uma língua gestual, proporcionando uma maior inclusão destes com o restante da sociedade.

Como contribuições adicionais, espera-se que a técnica de reconhecimento de LIBRAS possa ser utilizada por outros trabalhos, bem como o banco de sinais que está sendo organizado. Destaca-se que, devido aos desafios no reconhecimento de línguas de sinais, não espera-se que o sistema produzido seja uma solução definitiva para o reconhecimento de LIBRAS, mas sim uma ferramenta para auxiliar neste processo que poderá ser aperfeiçoada em trabalhos futuros. Por exemplo, com a incorporação de expressões não manuais.

4. Método de Pesquisa

A primeira etapa deste trabalho consiste de uma revisão sistemática da literatura sobre reconhecimento automático de línguas de sinais. Em seguida, com base nesta revisão, serão escolhidas as técnicas de segmentação e de reconhecimento de sinais que serão utilizadas na especificação e implementação de uma ferramenta para o reconhecimento automático de LIBRAS. Estas técnicas serão combinadas ou estendidas de forma a se construir um sistema robusto.

Em paralelo a etapa de implementação do sistema, pretende-se organizar um novo banco de imagens de sinais para ser utilizado nos testes e validação, sem a utilização de luvas coloridas e/ou laboratórios com ambientes controlados. O sistema será implementado, testado e validado comparando os resultados obtidos com os disponíveis na literatura e com a ajuda de uma especialista do domínio de LIBRAS. Para a validação, será medida a acurácia média da solução utilizando-se validação cruzada com divisão em 10 subconjuntos (*10-fold-cross-validation*).

5. Estado Atual do Trabalho

A pesquisa encontra-se em estado intermediário de desenvolvimento. A revisão sistemática da literatura já foi realizada e uma ferramenta para a identificação de sinais

dinâmicos também já foi especificada e desenvolvida. Para treinar e testar esta ferramenta foram utilizados inicialmente o banco de imagens produzidos na iniciação científica da aluna proponente (nos vídeos utilizados para formar este banco de imagens as pessoas sinalizantes utilizaram luvas coloridas para facilitar a segmentação) [Teodoro 2012]. Os resultados desta atividade foram publicados em [Teodoro and Digiampietri 2013]. Agora pretende-se aperfeiçoar a ferramenta implementada e testá-la utilizando o novo banco de imagens que está sendo construído, em que não são utilizadas luvas coloridas na gravação dos vídeos, diferentemente do banco de imagens utilizado inicialmente.

Além da construção do novo banco de imagens, também está em andamento a implementação de um algoritmo para a segmentação de imagens relacionadas à LIBRAS (sem a utilização de luvas coloridas). O segmentador já foi prototipado e obteve resultados promissores, mas ainda está sendo aprimorado a fim de atingir melhores resultados.

Agradecimentos

O trabalho apresentado neste artigo está sendo realizado com o auxílio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), por meio da concessão de Bolsa de Mestrado.

Referências

- Akmeliawati, R., Ooi, M.-L., and Kuang, Y. C. (2007). Real-time malaysian sign language translation using colour segmentation and neural network. In *Instrumentation and Measurement Technology Conference Proceedings (IMTC)*, pages 1–6.
- Albres, N. A. (2010). *Surdos & Inclusão Educacional*. Editora Arara Azul.
- Dias, J., Nande, P., Barata, N., and Correia, A. (2004). Ogre - open gestures recognition engine. In *Computer Graphics and Image Processing, 2004. Proceedings. 17th Brazilian Symposium on*, pages 33 – 40.
- Digiampietri, L. A., Teodoro, B., Santiago, C., Oliveira, G., and Araújo, J. (2012). Um sistema de informação extensível para o reconhecimento automático de libras. *SBSI 2012 - Trilhas Técnicas (Technical Tracks)*.
- Ferreira, A. L., Weck, J. T., da Silva, J. E. F., do Vale de Sousa, M. F., and de Andrade Santos, P. R. (2011). *Aprendendo Libras: módulo 2*. EDUFRN.
- Klima, E. and Bellugi, U. (1979). *The signs of language*. Cambridge University Press.
- Quadros, R. M. and Karnopp, L. B. (2004). *Língua de sinais brasileira - Estudos linguísticos*. Editora Artmed.
- Stokoe, W. C. (2005). Sign language structure: An outline of the visual communication systems of the american deaf. *Journal of Deaf Studies and Deaf Education*, 10(1):3–37.
- Teodoro, B. (2012). Desenvolvimento de ferramenta para análise de sequências de imagens e vídeos digitais em libras. *Trabalho de Conclusão de Curso, EACH-USP*.
- Teodoro, B. and Digiampietri, L. A. (2013). A local alignment based sign language recognition system. *Workshop of Works in Progress (WIP - SIBGRAPI), Arequipa, Peru*.