



INFORMAÇÕES GERAIS DO TRABALHO

Título do Trabalho:

Desenvolvimento de uma base de sinais da Língua Brasileira de Sinais utilizando gravações em 2D e 3D para avaliação de metodologias de reconhecimento de sinais em Visão Computacional.

Autor (es): Andreia Chagas Rocha Toffolo/Sílvia Grasiella Moreira Almeida

Palavras-chave: reconhecimento de sinais de Libras, base de sinais, elaboração de protocolo.

Campus: Ouro Preto

Área do Conhecimento (CNPq): Ciência da Computação/Letras

Tipo de bolsa: DTI-C

Financiador: Reitoria - IFMG

RESUMO

Apresenta-se, neste artigo, metodologia para a criação de uma base com sinais da Língua Brasileira de Sinais que possa ser utilizada em sistemas operacionais proprietários e livres. A etapa de criação descrita aqui tem por foco a definição de um protocolo consistente, que permitirá que sinais de Libras possam ser adicionados à base de forma normatizada. O protocolo deve versar sobre questões que envolvam a língua em si, a tecnologia utilizada para a gravação, e em técnicas que envolvam a própria gravação. Uma vez definido o protocolo, sinais em Libras serão gravados utilizando dois sensores RGB-D. Utilizaremos também uma câmera digital profissional baseada em sensor de imagem CMOS para gravar os sinais sob a mesma condição dos sensores RGB-D. Para operação dos sensores RGB-D, será utilizado o framework OpenNI, que é um SDK (*Software Development Kit*) de código aberto para desenvolver bibliotecas e aplicações de sensoriamento com dispositivos 3D. A importância desta base, quando pronta, é ser utilizada para testar a eficácia de sistemas computacionais baseados tanto em recursos de software livre quanto proprietários. A base com sinais de Libras e seu respectivo protocolo de criação ficarão disponíveis para uso de toda a comunidade acadêmica.

INTRODUÇÃO:

A linguagem é uma ferramenta poderosa que permite a interação do ser humano na sociedade das mais variadas formas, seja por sinais, códigos linguísticos, escrita e/ou oralidade. Utilizamos a língua para alcançar o principal fundamento da linguagem, a comunicação.

Há diversas classificações possíveis para as línguas existentes no mundo. De acordo com a classificação de línguas realizada em (LEWIS et al., 2009), existem 136 diferentes famílias de línguas chamadas de nível superior. Seis destas famílias destacam-se como as maiores famílias de línguas do mundo. Juntas, elas respondem por 2/3 de todas as línguas e atingem 5/6 da população mundial.

As outras 130 famílias incluem categorias especiais de línguas definidas como línguas construídas, línguas de sinais, línguas isoladas, línguas misturadas, línguas de contato e línguas ainda não classificadas. Em todos estes casos as línguas são utilizadas atualmente e, portanto, denominadas “vivas”.

O objeto de pesquisa deste trabalho é a Língua Brasileira de Sinais – Libras, que encontra-se na modalidade visual-espacial das línguas e é a língua oficial da comunidade surda no Brasil. Ao longo do tempo as abordagens educacionais com foco na aprendizagem dos surdos foram amadurecendo e sofrendo alterações de paradigmas fundamentais para a estruturação de uma língua própria que privilegiasse as especificidades dos surdos. A atual proposta de educação para surdos prevê um ensino bilíngue, em que a Libras é aceita e reconhecida como primeira língua (L1) e o português escrito como segunda língua.



A característica visual-motora da língua chama atenção especial da área de Visão Computacional, uma vez que o reconhecimento de sinais ainda é um grande desafio para essa área. O problema de reconhecimento automático de sinais em línguas diversas tem sido abordado pela área de Visão Computacional de forma bastante recorrente, como pode ser visto em ARAN et. al (2010), CARIDAKIS et. al (2012), LAHAMY et. al (2012) e em LY (2012).

Diversas técnicas desta área têm sido combinadas em propostas de abordagem para que um sinal, menor unidade possível da língua de sinais com significado, possa ser reconhecido de forma automática por sistemas computacionais. Assim, a diversidade das técnicas pode ser vista a partir dos trabalhos propostos em CARIDAKIS et. al (2012), COOPER et. al (2012), GEETHA et. al (2012) e KARAMI et. al (2011).

Entretanto, embora muitos tenham sido os sistemas automáticos desenvolvidos, observa-se nestes trabalhos a inexistência de bases de dados (chamadas *benchmarks*) que contenham imagens e/ou vídeos para testes, treinamento e validação de forma padronizada destes sistemas. Em geral, as pesquisas precisam ocupar-se de produzir a própria base de imagens e/ou vídeos, como acontece em PHADTARE et. al (2012), em AUTOR (2014), em REZENDE (2016). São diversos os motivos para que cada autor crie sua base. Entre os principais estão: (i) a tecnologia que envolve o trabalho em questão, (ii) a abordagem de alguma característica específica do vídeo, como face e/ou mãos, (iii) as versões dos *softwares* utilizados para operação dos sensores, (iv) a posição do sinalizador com relação aos sensores.

Diante das dificuldades em encontrar uma base com sinais da Libras, um protótipo de base de sinais foi criado em AUTOR (2014). Na construção desta base houve o cuidado de estabelecer especificações possíveis de serem reproduzidas para que mais sinais e sinalizadores pudessem ser acrescentados a ela. A base criada para o trabalho, entretanto, é considerada restrita pois contempla apenas um sinalizador, 34 sinais e cinco gravações de cada um deles. A referência para escolha dos 34 sinais deu-se a partir dos estudos descritos em CAPOVILLA et. al (2012a, 2012b).

O presente trabalho tem por objetivo criar um protocolo para construção de uma base de sinais, além de construir uma base com sinais de Libras a partir do protocolo definido. Esta base será composta por imagens de cor (chamadas aqui RGB), imagens de profundidade (Depth), informações de pontos do corpo humano e também vídeos gravados em formato 2D (linear, x e y) para uso pela comunidade científica em pesquisas de reconhecimento automático de sinais.

A Língua Brasileira de Sinais possui mais de dez mil verbetes, conforme documentado em CAPOVILLA et. al (2012a, 2012b). E, como toda língua viva, há uma mutação e expansão presentes de forma constante em sua estrutura. Assim, encontrar bases de sinais que possam ser utilizadas para testes em reconhecimento de sinais para uso da comunidade científica em Visão Computacional é, ainda, um desafio.

A relevância deste trabalho encontra-se na contribuição que pretende fornecer à área de reconhecimento de gestos (especificamente de sinais) em Visão Computacional, além da relevância na área social-cultural para a comunidade surda, uma vez que se propõe a criar meios para tornar mais consistentes as pesquisas em tecnologias assistivas para surdos.

Ao construir-se uma base de sinais em Libras de forma sistematizada, as pesquisas por técnicas que permitam sistemas de reconhecimento automático de sinais com altas taxas de reconhecimento terão uma opção de *benchmark* para teste e treinamento importante, pois comparações entre sistemas utilizando a mesma base tornam estes sistemas mais robustos.

Tal pesquisa também vem de encontro às políticas de inclusão social, na medida em que sistemas deste tipo permitem maior divulgação da língua de sinais e meios tecnológicos para que sua aprendizagem possa possuir mais um instrumento.

Ademais, os sistemas automáticos podem ser utilizados em ambientes e/ou momentos nos quais as pessoas que desejam e/ou precisam comunicar-se sejam falantes de Libras e Português, mas não conheçam a língua um do outro.



METODOLOGIA:

A etapa atual do trabalho está na definição de um protocolo de criação de bases de sinais. As seguintes definições já foram estabelecidas:

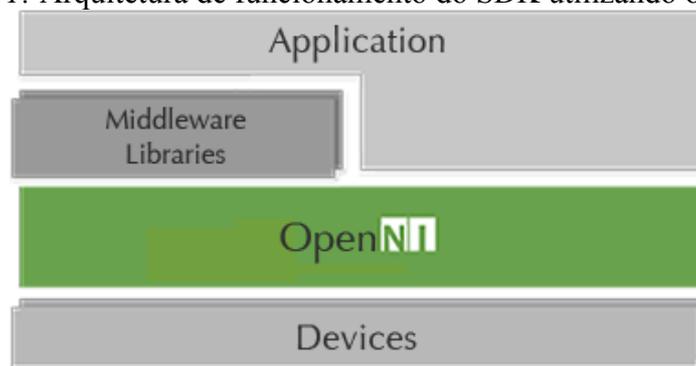
- (i) as gravações serão realizadas por dois sensores RGB-D Kinect e uma máquina digital Canon EOS Rebel t5i;
- (ii) será gravado todo o movimento corporal que envolve a sinalização, englobando a leitura da expressão facial e de parâmetros relacionados ao movimento e posição das mãos;
- (iii) será utilizado o *framework* OpenNI SDK como *software* de operação dos sensores RGB-D;
- (iv) o sinalizador estará em frente aos sensores na hora da gravação;
- (v) serão 20 sinalizadores, entre homens e mulheres, com características de altura, peso, cor diferenciadas.

Já as questões que estão em estudo para definição do protocolo são:

- (i) número de sinais que serão gravados, a partir de um critério de escolha destes;
- (ii) quais serão os sinais gravados;
- (iii) qual será o ponto inicial e o ponto final da gravação de cada um dos sinais;
- (iv) qual a cor de fundo das gravações;
- (v) se haverá padrão para a vestimenta;
- (vi) qual será o padrão de luminosidade, distância do sinalizador ao sensor, resolução da imagem e posição de captura dos movimentos.

O *framework* OpenNI é um kit de desenvolvimento de sistemas de código aberto usado para o desenvolvimento de bibliotecas e aplicações de sensoriamento 3D.

Figura 1: Arquitetura de funcionamento do SDK utilizando o OpenNI



Fonte: <http://www.openni.ru/>, 2017.

Como referência com relação a protocolos utilizados em bases de vídeos com movimento, podemos citar o trabalho apresentado em Zhang (2016), que apresenta uma comparação entre 26 bases de ações e atividades realizadas em vista única com relação a: (i) número de ações, sujeitos e amostras; (ii) formato dos dados (Cor, Profundidade, Esqueleto, Máscara humana, Nuvem de pontos, dados do sensor; (iii) notação para o protocolo (sujeito cruzado, realização de uma sequência, validação cruzada). Nenhuma destas bases é específica para sinais de Libras.

A escolha dos sinais que serão gravados na primeira versão da base de sinais é de fundamental importância porque vai conduzir outras possíveis gravações e deve permitir a maior diversidade de situações possível. Portanto, esta escolha é que tornará a base robusta e não polarizada para um único tipo de sinal.



Decidiu-se, assim, utilizar como referência para escolha daqueles que constituirão nossa base, sinais que estão disponíveis no projeto *Identificador de Sinais*, desenvolvido em Quadros (2017), pesquisadora da Universidade Federal de Santa Catarina (UFSC). A UFSC é pioneira na implementação do curso de Letras-Libras, cujo objetivo é a formação de professores para atuar no ensino da língua de sinais como primeira e segunda língua. Tal instituição destaca-se pelo investimento em ações voltadas à comunidade surda além de contar com profissionais renomados na área, o que nos motivou a trabalhar com os sinais disponibilizados por essa equipe. O *Identificador de Sinais* tem como principal objetivo organizar e disponibilizar aos pesquisadores acesso às glosas¹ padronizadas pelos transcritores do Núcleo de Aquisição de Língua de Sinais/NALS da UFSC, para realização da transcrição de dados em Libras para português.

A busca de sinais pode ser realizada pelo nome em português, inglês ou pelo sinal (configuração de mão ou localização). Atualmente o *ID Sinais* é composto por mil sinais. Em nossa base, selecionaremos sinais a partir da informação de localização em que estes são realizados, sejam elas: neutra, olhos, nariz, boca, orelhas, ombros e braços.

RESULTADOS E DISCUSSÕES:

Como resultado, já foram estabelecidas algumas questões técnicas, trabalho que foi realizado em conjunto com o Laboratório de Inteligência Computacional MINDS, da UFMG., Departamento de Engenharia Elétrica. Segue o que foi definido até o momento como protocolo para a criação da base de sinais:

- Número de gestos: 104, conforme distribuição a seguir. Eles serão selecionados a partir de e com autorização da equipe do IDSinais, da UFSC.

Localização	Número de sinais da base do IDSinais	%Qtde sinais gravados (10%)
Olhos e nariz	126	13
Boca e queixo	191	19
pescoço	48	5
Ombro e Peito	72	7
Braço	96	10
Neutro	1141*	50
TOTAL DE SINAIS	1674	104

*Como a quantidade de sinais é grande se comparado aos outros, selecionaremos cerca de 4%.

- Número de pessoas: 10 homens e 10 mulheres.
- Número de exemplos: 5 gravações de cada pessoa.
- Tipo: single view
- Sensores: 2 sensores RGB-D (Kinect 2010 e Kinect 2014) e uma câmera CMOS.
- Dados capturados:
 - Pelos sensores RGB-D: cores (RGB) com baixa resolução e com alta resolução, depth, skeleton.
 - Pela câmera CMOS: RGB (alta resolução).

¹ Glosa é o nome que identifica o sinal, a identidade do sinal. Alguns sinais tem diferentes possibilidades de tradução de acordo com o contexto. Diferente do nome que identifica o sinal (identidade do sinal) que não se altera mesmo que ocorra em diferentes contextos. Por exemplo, o sinal pode ser traduzido como GOSTAR na variante dos sinalizantes do Rio Grande do Sul, no entanto para atender os objetivos da padronização de transcrição será sempre identificado como VONTADE que representa a palavra mais comum de ser associada ao sinal apresentado.
<http://idsinais.libras.ufsc.br/perguntasfrequentes.php>



É importante a criação de uma base de sinais em Libras para impulsionar trabalhos realizados em reconhecimento automático de sinais. Entretanto, neste momento o mais relevante é permitir que futuramente sinais possam ser incorporados a ela por meio da mesma técnica e sob as mesmas condições. E é por isto que a definição do protocolo de criação é o foco deste trabalho, pois todo o procedimento que envolva a construção da base deve ser controlado e documentado de forma bastante detalhada.

O presente projeto é interdisciplinar, envolvendo profissionais de três instituições federais de ensino e seus respectivos grupos de trabalho, além de pesquisadores das áreas de reconhecimento de padrões, inteligência computacional e língua brasileira de sinais.

Por fim, entendemos que nosso objetivo é viável e será realmente de grande importância em etapas de simulação, validação e testes em produtos da área de tecnologia assistiva em Língua Brasileira de Sinais por meio de Visão Computacional.

REFERÊNCIAS BIBLIOGRÁFICAS:

1. [Autor], Tese de Doutorado, Universidade Federal de Minas Gerais, 2014.
2. ARAN, O.; AKARUN, L. *A multi-class classification strategy for Fisher scores: Application to signer independent sign language recognition*. Pattern Recognition, v. 43, n. 5, p. 1776–1788, 2010.
3. CAPOVILLA, F. C.; RAPHAEL, W. D.; MAURÍCIO, A. C. L. *Novo Deit-Libras: Dicionário Enciclopédico Ilustrado Trilíngue da Língua Brasileira de Sinais (Libras) baseado em Linguística e Neurociências Cognitivas*, Volume I: Sinais de A a H. [S.l.]: Edusp, 2012.
4. CAPOVILLA, F. C.; RAPHAEL, W. D.; MAURÍCIO, A. C. L. *Novo Deit-Libras: Dicionário Enciclopédico Ilustrado Trilíngue da Língua Brasileira de Sinais (Libras) baseado em Linguística e Neurociências Cognitivas*, Volume II: Sinais de I a Z. [S.l.]: Edusp, 2012.
5. CARIDAKIS, G. et al. *Non parametric, self-organizing, scalable modeling of spatiotemporal inputs: The sign language paradigm*. Neural Networks, v. 36, p. 157–166, 2012.
6. GEETHA, M.; MANJUSHA, U. C. *A vision based recognition of Indian Sign Language alphabets and numerals using B-spline approximation*. International Journal on Computer Science & Engineering, v. 4, n. 3, 2012.
7. KARAMI, A.; ZANJ, B.; SARKALEH, A. K. *Persian Sign Language (PSL) recognition using Wavelet Transform and Neural Networks*. Expert Systems with Applications, v. 38, n. 3, p. 2661–2667, 2011.
8. LAHAMY, H.; LICHTI, D. D. *Towards real-time and rotation-invariant American Sign Language alphabet recognition using a range camera*. Sensors, v. 12, n. 11, 2012.
9. LEWIS, M. P. et al. *Ethnologue: Languages of the world*. [S.l.]: SIL international Dallas, TX, 2009.
10. LI, Y. *Hand gesture recognition using Kinect*. In: 2012, IEEE 3rd International Conference on Software Engineering and Service Science. [S.l.: s.n.], 2012. (ICSESS 2012), p. 196–199.
11. PHADTARE, L.; KUSHALNAGAR, R.; CAHILL, N. *Detecting hand-palm orientation and hand shapes for sign language gesture recognition using 3D images*. In: Western New York Image Processing Workshop. [S.l.: s.n.], 2012. (WNYIPW 2012), p. 29–32.
12. REZENDE, T.M. *Aplicação de Técnicas de Inteligência Computacional para Análise da Expressão Facial em Reconhecimento de Sinais de Libras*. Dissertação de Mestrado. Mestrado, Universidade Federal de Minas Gerais, 2016.



13. QUADROS, R. M.; OLIVEIRA, J.S.; SANTOS, P.R.C.; MIRANDA, R.D.; *ID - Identificador de Sinais*. Disponível em <http://www.idsinais.libras.ufsc.br/index.php>. Acesso em Fevereiro de 2017.

Participação em Congressos, publicações e/ou pedidos de proteção intelectual:

XIV EVIDOSOL e XI CILTEC-Online - junho/2017 - <http://evidosol.textolivre.org>