
Interface para Reconhecimento da Língua Brasileira de Sinais

Paulo Marcotti^{1,3}, Luciana Babberg Abiuzi¹, Paloma Maria Silva Rocha Rizol^{1,2},
Carlos Henrique Quartucci Forster¹

¹Divisão de Computação – Instituto Tecnológico de Aeronáutica (ITA)
CEP 12.228-900 – São Jose dos Campos – SP – Brasil

²Universidade Estadual Paulista (UNESP)
Campus de Guaratinguetá – SP – Brasil

³Faculdade de Tecnologia Termomecânica (FTT)
São Bernardo do Campo – SP – Brasil

forster.ita.br, abiuzi@ita.br, paloma@feg.unesp.br,
marcotti@comp.ita.br

Abstract. *Computers are useful in several areas, as well as the use of computer science in the education and the software development for communication is increasing interaction between people. This paper presents some computer vision techniques applied in a computer-mediate communication system model, which uses an image classifier to recognize sign language (Brazilian Sign Language) in gestural animation form and show to user the respective word in Portuguese.*

Resumo. *A popularização do computador e sua utilização em diversas áreas são fatos inquestionáveis, assim como o uso da informática na educação e o crescente desenvolvimento de softwares educacionais e para comunicação e interatividade entre pessoas. Este artigo apresenta um conjunto de técnicas da área de Visão Computacional aplicado em um modelo de sistema de comunicação mediado por computador: trata-se de um classificador de imagens capaz de reconhecer determinados sinais da Língua Brasileira de Sinais e apresentar ao usuário o respectivo significado em Português.*

1. Introdução

A Libras – Língua Brasileira de Sinais – é reconhecida como meio legal de comunicação e expressão entre as comunidades de pessoas surdas no Brasil. Ganhou esse status porque surgiu naturalmente assim como a língua portuguesa e atualmente é de grande importância na comunicação no território brasileiro. A Libras baseou-se primariamente na Língua de Sinais Francesa, apresentando semelhanças em relação a várias línguas de sinais européias e à norte-americana. Assim como as diversas línguas existentes, ela é composta por níveis lingüísticos como: fonologia, morfologia, sintaxe e semântica. Com o avanço da Internet, faz-se necessário o uso de ferramentas que auxiliem a comunicação entre surdos e pessoas que não conhecem os sinais da Libras, assim este trabalho visa possibilitar ou aumentar a comunicação entre a comunidade dos surdos e ouvintes utilizando o computador como um mediador.

O censo do IBGE do ano 2000 verificou que no Brasil existiam 5.750.000 pessoas com deficiência auditiva, sendo que, mais de 406.000 encontravam-se em idade escolar. Já o censo escolar do ano 2003, realizado pelo INEP, apontou que apenas 56.024 pessoas surdas estavam matriculadas em escolas do Ensino Fundamental e Médio. Destas pessoas, somente 2.041 conseguiram concluir o Ensino Médio nesse mesmo ano e, somente 344 foram identificadas como alunos matriculados nas universidades brasileiras [IBGE, INEP].

Embora exista uma iniciativa do Governo Brasileiro em prover condições para a educação dos alunos que apresentam surdez, uma das maiores dificuldades encontradas atualmente que impacta no processo de aprendizado está na indisponibilidade de recursos que facilitam a interação professor-aluno. O não conhecimento da Libras também é um dos fatores que influencia no baixo índice de alunos matriculados nas escolas. Portanto, dentro deste panorama, é perceptível a necessidade da existência de ferramentas computacionais que auxiliem o aprendizado dos deficientes auditivos e, o mais importante, que tenham efeito de sociabilidade, visando diminuir a exclusão social e digital dos mesmos.

O objetivo deste trabalho é demonstrar as etapas do processo de implementação de um software classificador de imagens, denominado *ClassLib*, cuja finalidade é a de traduzir os sinais da Libras para português.

2. Arquitetura do Sistema

Pode-se dizer que a classificação é um processo de extração de informação em imagens para reconhecimento de padrões e objetos similares.

As técnicas computacionais abordadas neste trabalho estão voltadas para a análise e processamento de imagens, que se dividem nas seguintes fases: aquisição de imagem, pré-processamento, segmentação, extração de atributos, reconhecimento e interpretação [Gonzalez 1992].

No atual estágio desta pesquisa, as técnicas mais exploradas foram aquelas que envolveram as etapas de extração de características, reconhecimento e interpretação de imagem.

2.1. Aquisição de Imagens

Na etapa de aquisição de imagem, optou-se pela coleta de quatro conjuntos distintos de imagens, cada qual contendo 21 fotos de sinais diferentes da Libras. Os sinais escolhidos foram:

Letras: B, C, L, M, P, Q e V.

Palavras: Amigo, Aranha, Borboleta, Casa, Chorar, Chuva, Criança, Dormir, Desculpas, Estudar, Livro, Macaco, Pensar, e Telefonar.

Número: Hum.

Três conjuntos de fotos foram utilizados para o treinamento do classificador, ou seja, para a extração de características, e o último conjunto foi utilizado para a fase de testes do classificador. Embora estes sinais sejam uma quantidade muito reduzida do conjunto de todos os sinais da Libras, o foco deste artigo é validar que características

simples de serem calculadas podem sub-dividir o conjunto de todos os sinais da Libras em sub-conjuntos menores e que podem ser diferenciados posteriormente com outra técnica de tomada de decisão.

2.2. Pré-Processamento

Após a etapa de aquisição, os conjuntos de imagens foram submetidos a um pré-processamento manual. No primeiro conjunto de fotos coletado, o fundo de todas as imagens foi editado para a cor preta, todas foram convertidas para o formato *bitmap* e para o tamanho 320x240 *pixels*. Nos três últimos conjuntos, o formato e a dimensão das figuras foram mantidos de forma compatível ao primeiro conjunto, porém, o fundo das imagens não foi alterado.

Após uma prévia segmentação da imagem, etapa esta a ser explicada no próximo tópico, foi aplicada a técnica de binarização na região onde se encontra a mão a fim de remover ruídos e possibilitar a extração de características do sinal em questão.

A técnica de binarização pode ser utilizada quando se tem interesse em obter resolução de formas (contornos, padrões geométricos específicos, entre outros) em imagens. Tal procedimento consiste em converter uma imagem para apenas dois valores, expressados por símbolos binários: branco e preto.

Neste trabalho, a binarização foi realizada pelo software *ClassLib* de forma simplificada, onde cada *pixel* da área do sinal foi alterado para branco ou preto avaliando um limiar de intensidade $I = (R + G + B) / 3$. Porém, esta técnica pode apresentar limitações se o gesto estiver próximo a áreas claras, como face, fundo branco ou reflexo de flash, pois, mesmo que padronizadas, as métricas de cores sofrem variações devido à luminosidade do ambiente e posicionamento de fontes luminosas [Swenson 2000].



Figura 1. Pré-processamento manual das imagens.

2.3. Segmentação

Pode-se dizer que a fase de segmentação é responsável por particionar a imagem em regiões disjuntas que devem corresponder às áreas de interesse da aplicação. A partir das áreas segmentadas são extraídas as características utilizadas para classificar as imagens.

Nas imagens, a segmentação delimitou-se em duas fronteiras: ‘área útil’ e ‘área de sinal’ da imagem. A área útil é aquela onde se encontra a pessoa na foto e a área de sinal é aquela onde se encontra a área da mão com respectivo gesto da Libras. A segmentação da imagem é realizada manualmente através do software *ClassLib*.

Na delimitação da ‘área do sinal’ da imagem foi aplicada a técnica de binarização de imagem, conforme explicado no tópico anterior, onde, o sinal (mão) foi transformado na cor branca e o fundo da área do sinal foi transformado na cor preta.

2.4. Extração de Características

Nesta fase do processo, efetuou-se a extração de características das áreas segmentadas dos três conjuntos de imagens utilizados para treinamento do classificador. Para cada sinal foram calculadas as seguintes características:

- a) Posição relativa no plano da imagem, representada pelo centro de gravidade do objeto (mãos) nos eixos horizontal e vertical. O valor da posição é dividido pela largura e altura da ‘área útil’ segmentada, correspondendo a valores relativos do ponto na imagem;
- b) Área relativa do objeto é a área da mão fazendo o gesto em relação à ‘área útil’ segmentada. Esta característica foi calculada de forma relativa para eliminar o efeito de fotos próximas ou fotos distantes;
- c) Momento de inércia em torno dos eixos horizontal e vertical, com a finalidade de medir a inclinação do objeto. Esta característica também é calculada de forma adimensional, ou seja, é dividida pela ‘área do sinal’ segmentada, evitando o efeito da proximidade entre a câmera e o objeto.

Os sinais, mesmo que executados por intérpretes diferentes em ambientes diferentes, têm características em comum. Por exemplo, o sinal da palavra ‘amigo’ em Libras é sempre feito na altura do peito esquerdo da pessoa, com a mão esticada e de lado, deste modo o centro de gravidade que caracteriza a posição da mão é idêntico e com inclinação similar em todas as fotos, considerando-se o ponto relativo ao corpo da pessoa.

É possível que outras características sejam necessárias, como direção da abertura da mão, posição relativa entre as mãos, ou também relação da área da envoltória convexa que pode representar uma medida se a mão está ‘espalmada’ ou ‘fechada’. Uma característica primordial, a ser abordada na próxima versão do *ClassLib*, é a expressão facial do intérprete, que é vital para o correto entendimento de muitos sinais parecidos.

A extração das características utilizadas para treinamento foi realizada pelo software *ClassLib*. Posteriormente, tais características foram utilizadas pelo software *Weka* para a geração de uma árvore de decisão utilizada para classificar imagens.

2.5. Reconhecimento e Interpretação

O reconhecimento é o processo que atribui um rótulo a um objeto, baseado em suas características. A interpretação envolve a atribuição de significado a um conjunto de objetos reconhecidos. O reconhecimento e a interpretação de um sinal baseiam-se em características extraídas da imagem [Ramesh 1995].

Existem várias técnicas para classificação que podem ser utilizadas nessa fase. Foi realizado um estudo sobre ‘árvore de decisão’, que é uma ferramenta utilizada na área de Inteligência Artificial para identificação de padrões e para separar em classes objetos de acordo com valores de suas características. Uma Árvore de Decisão é fácil de

ser implementada em um software por ser pequena e se expande apenas o suficiente para conseguir classificar o conjunto de exemplos de treinamento disponível, mesmo considerando o ruído nos dados. Os nós mais próximos de sua raiz contêm os atributos que oferecem o maior ganho de informação, desse modo permitem decidir rapidamente a que subconjunto pertence.

Após gerar um conjunto de características para cada imagem pertencente ao grupo de treinamento, tais características foram exportadas, através do software *ClassLib*, para uma planilha *MSEXcel*. Posteriormente, esta planilha foi importada no software *Weka*, para a geração de uma árvore de decisão capaz de classificar os sinais treinados através de suas características.

No software *Weka*, foram realizados vários testes com diversos classificadores, a fim de encontrar uma árvore de decisão adequada que pudesse ser implementada no software *ClassLib*. Optou-se pela implementação do algoritmo 'J48', onde a árvore gerada pelo *Weka* é demonstrada na Fig. 2.

Após a implementação da árvore de decisão, foi apresentado ao software *ClassLib* um sinal do grupo de imagens não utilizado para treinamento, de modo que as características da imagem em questão foram calculadas e submetidas para as condições da tabela de decisão, a fim de resultar um único sinal da Libras.

A grande variabilidade nos sinais entre intérpretes diferentes ou o mesmo sinal feito em momentos diferentes pela mesma pessoa provocou algumas discrepâncias em relação à classificação de determinados sinais. Essa aparente dificuldade indica e estimula novos estágios da pesquisa de reconhecimento e interpretação de imagens. Os sinais da Libras são numerosos, aproximadamente 8000 sinais, e para facilitar esta etapa de reconhecimento e interpretação precisam ser 'clusterizados', utilizando as mesmas características, e em seguida deve ser realizada uma avaliação mais criteriosa para serem separados em pequenos grupos de sinais.

3. Desenvolvimento de Interface Gráfica

Neste capítulo, tem-se a intenção de demonstrar o processo de criação de interface do software *ClassLib*. Podemos dividir o processo de criação de interface em três partes: navegação, entrada de dados e saída de dados [Dennis 2005].

Em relação ao processo de navegação, procurou-se estabelecer um padrão no desenvolvimento das telas, na utilização dos ícones que compõem os botões, na geração dos relatórios e nas mensagens exibidas eventualmente aos usuários. No processo de entrada de dados, procurou-se facilitar a entrada das informações no sistema, evitando a poluição de componentes na tela e tornado o processo de inclusão, alteração e classificação de dados de fácil compreensão. No processo de saída dos dados, o objetivo foi o de permitir ao usuário a geração de informação de forma consistente, precisa e de mínimo esforço para compreensão [Oliveira 2004].

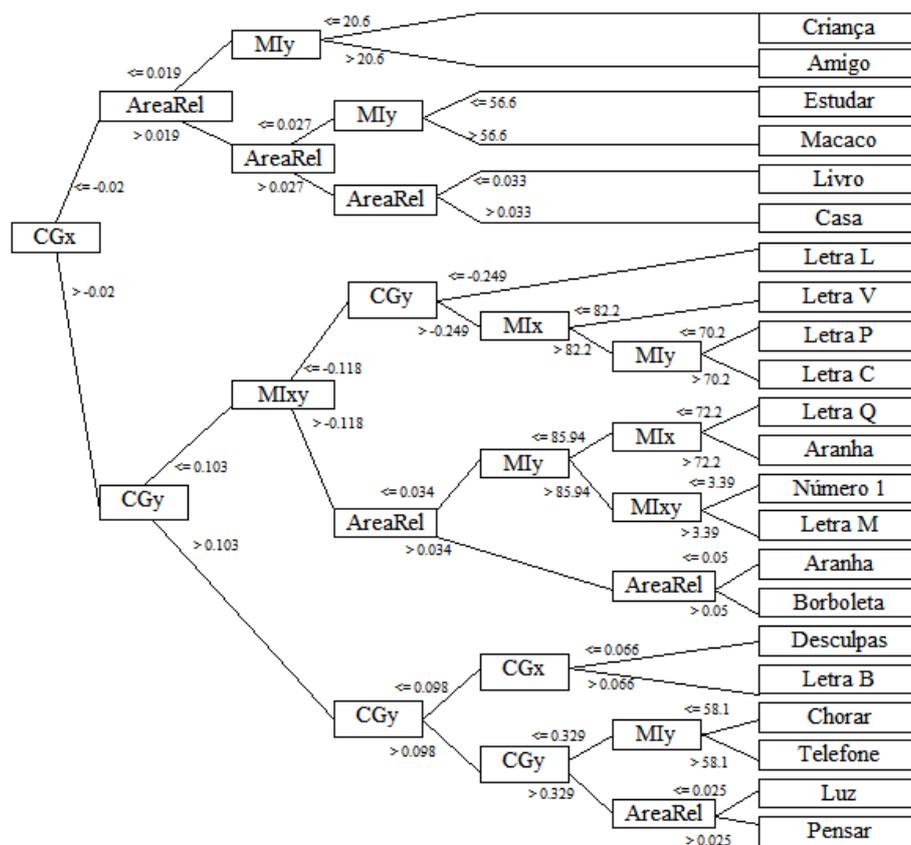


Figura 2. Árvore de decisão gerada no software Weka. As variáveis utilizadas possuem os seguintes significados: CGx e XGy são os centros de gravidade na direção dos eixos x e y em relação ao centro da área do sinal; Mlx, Mly e Mlxy são os momentos de inércia de inclinação da área do sinal e a AreaRel é a área da mão dividida pela área útil do corpo da pessoa.

A tela de treinamento (Fig. 3) é considerada um local de entrada de dados no sistema. É a partir desta tela que as imagens são carregadas, processadas e armazenadas, juntamente com as suas características, na base de dados. Ao incluir ou alterar um registro, o programa permite que uma imagem seja carregada e vinculada à sua correspondente descrição do sinal de Libras. Neste processo é permitida ao usuário a seleção das áreas ‘útil’ e ‘de sinal’, que são utilizadas para calcular os atributos da imagem, conforme descrito anteriormente no capítulo 2.

A partir das imagens e suas respectivas características armazenadas na base de dados através da tela de treinamento, os dados do *ClassLib* foram exportados para o software *MSExcels* e submetidos a um treinamento de classificação no software *Weka*. O algoritmo obtido no software *Weka* foi implementado na tela de classificação a fim de possibilitar a identificação das classes das imagens carregadas e processadas nesta tela.

A tela de classificação (Fig 5) possui os mesmos princípios da tela de treinamento. É utilizada para carregar uma imagem, permite a seleção das áreas ‘útil’ e ‘de sinal’ desta imagem e efetua o cálculo das suas características. A diferença consiste na classificação da imagem, onde as características desta são submetidas a um algoritmo que verifica e demonstra ao usuário à qual classe a imagem pertence.

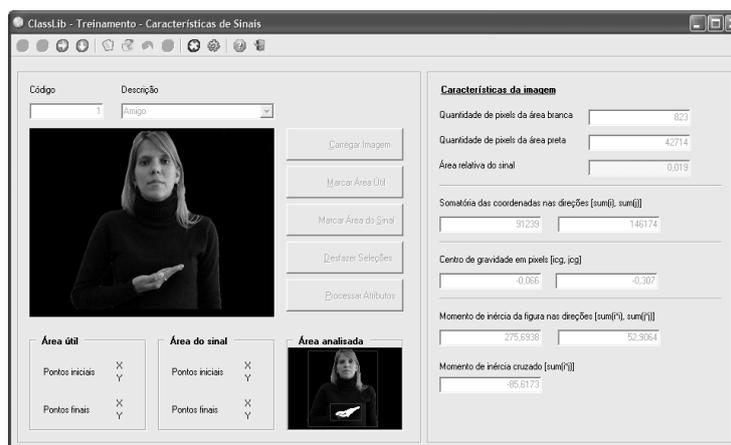


Figura 3. Tela de treinamento do *ClassLib*, em modo de visualização.

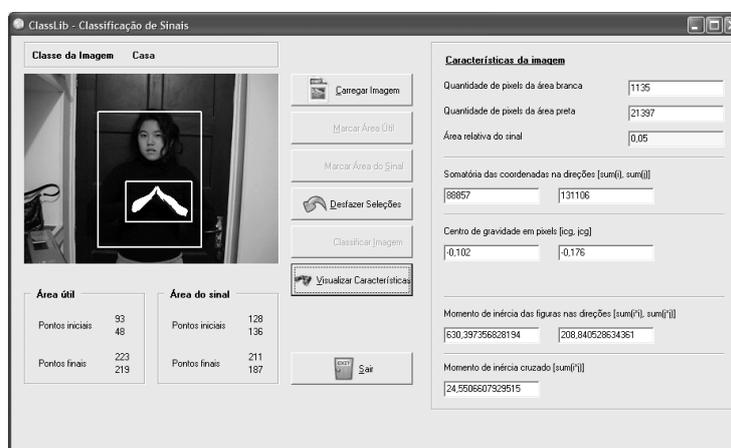


Figura 4. Tela de classificação do *ClassLib*, com as características das imagens sendo exibidas.

4. Conclusão e Trabalhos Futuros

De uma forma geral o sistema implementado apresentou boa interface gráfica com o usuário. Porém, este protótipo ainda apresenta limitações, que serão objeto da evolução deste projeto. O sistema pode ser muito útil no ensino a distância de pessoas surdas, e que também pode auxiliar alunos e professores no aprendizado da Libras.

As características extraídas e utilizadas se mostraram eficientes. O reconhecimento deve se dividir em 2 fases: uma para encontrar qual o grupo a que pertence o sinal da Libras e outra uma busca refinada dentro de um subconjunto restrito de sinais. Para a primeira etapa do reconhecimento estas características devem ser mantidas devido à sua facilidade de extração.

Um ferramenta como o *ClassLib* pode ser considerada um passo inicial para a implementação de um tradutor de Libras para Português, utilizando as teorias da ‘Comunicação Mediada por Computador’. A realidade do surdo brasileiro impôs que ele criasse condições para entender o mundo dos ouvintes, e é bastante comum o surdo conhecer a Libras e também ter a habilidade na leitura labial; também existem ferramentas computacionais para traduzir de Português para Libras (embora palavra por palavra). O presente estudo pretende propiciar um canal de comunicação que está

faltando: os ouvintes conseguem entender os sinais da Libras. Os trabalhos iniciais realizados por este grupo de pesquisas devem abrir caminho para uma ferramenta eficiente de comunicação, permitindo que o canal na direção do surdo para o ouvinte, e desse modo permita o perfeito entendimento entre aluno e professor.

As conclusões deste trabalho encorajam a evolução do presente estudo. Pelo caráter social deste trabalho as ferramentas deverão ser disponibilizadas em código aberto, e eventuais patentes deverão resguardar os interesses das comunidades dos surdos e dos que tenham poucos recursos financeiros e tecnológicos a sua disposição.

Referências

- Dennis, A., Wixowm, B.H.: Análise e Projeto de Sistemas. Rio de Janeiro: LTC Editora, (2005)
- Federação Nacional de Educação e Integração dos Surdos, <http://www.feneis.com.br>
- Gonzalez, R.C., Woods, R.E.: Digital image processing. New York: Addison Wesley (1992)
- Oliveira Netto, A.A.: IHC: Modelagem e Gerência de Interfaces com o Usuário. Visual Books: Florianópolis (2004)
- Ramesh, J., Rangachar, K., Schunk, B.: Machine Vision. McGraw-Hill, New York (1995)
- Reconhecimento da Língua Brasileira de Sinais – Libras como meio legal de comunicação e expressão, <http://www.planalto.gov.br/ccivil/LEIS/2002/110436.htm>
- IBGE – Instituto Brasileiro de Geografia e Estatística, <http://www.ibge.gov.br>
- INEP – Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira, <http://www.inep.gov.br>
- Weka University of Waikato, <http://www.weka.com>
- Swenson, R.L.: A Real-Time High Performance Universal Color Transformation Hardware System. University of Kent at Canterbury: Canterbury-Kent (2000)