FERNANDO DE ALMEIDA FREITAS

Reconhecimento automático de expressões faciais gramaticais na língua brasileira de sinais

São Paulo

2015

FERNANDO DE ALMEIDA FREITAS

Reconhecimento automático de expressões faciais gramaticais na língua brasileira de sinais

Versão corrigida

Versão corrigida contendo as alterações solicitadas pela comissão julgadora em 16 de março de 2015. A versão original encontra-se em acervo reservado na Biblioteca da EACH-USP e na Biblioteca Digital de Teses e Dissertações da USP (BDTD), de acordo com a Resolução CoPGr 6018, de 13 de outubro de 2011.

Orientador: PROFA. DRA. SARAJANE MARQUES PERES

São Paulo

2015

CATALOGAÇÃO-NA-PUBLICAÇÃO
(Universidade de São Paulo. Escola de Artes, Ciências e Humanidades Biblioteca)

Freitas, Fernando de Almeida
Reconhecimento automático de expressões faciais gramaticais na língua brasileira de sinais / Fernando de Almeida Freitas ; orientador, Sarajane Marques Peres. – São Paulo, 2015
112 f. : il.

Dissertação (Mestrado em Ciências) - Programa de Pós-Graduação em Sistemas de Informação, Escola de Artes, Ciências e Humanidades, Universidade de São Paulo
Versão corrigida

1. Aprendizado computacional. 2. Reconhecimento de padrões. 3. Interface homem-computador. 4. Língua Brasileiora de Sinais. 5. Análise do movimento humano. 6. Gestos – Análise. I. Peres, Sarajane Marques, orient. II. Título.

CDD 22.ed. – 006.31

Dissertação de autoria de Fernando de Almeida Freitas, sob o título **"Reconhecimento automático de expressões faciais gramaticais na língua brasileira de sinais"**, apresentada à Escola de Artes, Ciências e Humanidades da Universidade de São Paulo, para obtenção do título de Mestre em Ciências pelo Programa de Pós-graduação em Sistemas de Informação, na área de concentração Sistemas de Informação, aprovada em 16 de março de 2015 pela comissão julgadora constituída pelos doutores:

Profa. Dra. Sarajane Marques Peres
Presidente
Instituição: Universidade de São Paulo

Prof. Dr. José Mario de Martino
Instituição: Universidade Estadual de Campinas

Prof. Dr. Luciano Antonio Digiampietri
Instituição: Universidade de São Paulo

Dedico esta dissertação de Mestrado à comunidade surda, responsável pela criação destas expressões faciais que dizem mais que palavras...

Agradecimentos

Primeiramente agradeço a Deus, pelo seu amor incondicional e direcionamento em minha vida.

Agradeço aos meus pais e familiares pelo carinho e incentivo para chegar até aqui.

Agradeço em especial à minha orientadora Dra. Sarajane Marques Peres por toda dedicação e por ser um exemplo como pessoa para mim e ter me incentivado e direcionado durante toda esta dissertação.

Agradeço em especial também ao Dr. Felipe Venâncio Barbosa por ter colaborado durante o desenvolvimento deste trabalho e pelo amigo que se tornou.

Agradeço à minha namorada, Eloize, pelo carinho, paciência e incentivo durante o período do mestrado.

Por fim, agradeço aos professores do PPgSI pela oportunidade de receber um ensino com qualidade e profundidade.

"...se necessário for, use palavras."

(Franciso de Assis)

Resumo

FREITAS, Fernando de Almeida. **Reconhecimento automático de expressões faciais gramaticais na língua brasileira de sinais**. 2015. 112 f. Dissertação (Mestrado em Ciências) – Escola de Artes, Ciências e Humanidades, Universidade de São Paulo, São Paulo, 2015.

O reconhecimento das expressões faciais tem atraído bastante a atenção dos pesquisadores nas últimas décadas, principalmente devido às suas ponteciais aplicações. Nas línguas de sinais, por serem línguas de modalidade visual-espacial e não contarem com o suporte sonoro da entonação, as expressões faciais ganham uma importância ainda maior, pois colaboram também para formar a estrutura gramatical da língua. Tais expressões são chamadas expressões faciais gramaticais e estão presentes nos níveis morfológico e sintático das línguas de sinais. Elas ganham destaque no processo de reconhecimento automático das línguas de sinais pois colaboram para retirada de ambiguidades entre sinais que possuem parâmetros semelhantes, como configuração de mãos e ponto de articulação, além de colaborarem na composição do sentido semântico das sentenças. Assim, esta dissertação de mestrado tem por objetivo desenvolver um conjunto de modelos de reconhecimento de padrões capazes de resolver o problema de reconhecimento automático de expressões faciais gramaticais, usadas no contexto da Língua Brasileira de Sinais (Libras), considerando-as em nível sintático.

Palavras-chaves: Expressões Faciais Gramaticais. Reconhecimento de Padrões. Língua Brasileira de Sinais. Língua de Sinais.

Abstract

FREITAS, Fernando de Almeida. **Automatic recognition of Grammatical Facial Expressions from Brazilian Sign Language (Libras)**. 2015. 112 p. Dissertation (Master of Science) – School of Arts, Sciences and Humanities, University of São Paulo, São Paulo, 2015.

The facial expression recognition has attracted most of the researchers attention over the last years, because of that it can be very useful in many applications. The sign language is a spatio-visual language and it does not have the speech intonation support, so facial expression gain relative importance to convey grammatical information in a signed sentence and they contributed to morphological and/or syntactic level to a sign language. Those expressions are called grammatical facial expression and they cooperate to solve the ambiguity between signs and give meaning to sentences. Thus, this master thesis aims to develop models that make possible to recognize automatically grammatical facial expressions from Brazilian Sign Language (Libras).

Keywords: Grammatical Facial Expressions. Pattern Recognition. Brazilian Sign Language. Sign Language.

Lista de figuras

Lista de algoritmos

Lista de tabelas

Sumário

[1] De acordo com a Associação Brasileira de Normas Técnicas. NBR 6023.

1 Introdução

O reconhecimento automatizado das expressões faciais (EFs) atrai a atenção de pesquisadores principalmente devido ao seu potencial de aplicação, como no desenvolvimento de aplicações que possam fazer uso de análise de emoções expressas durante um discurso, durante a realização de compras ou em diversas outras ações; ou na síntese de emoções em animações e robôs; ou ainda em sistemas que usam análise de comportamento humano para monitoramento de pessoas, buscando garantir segurança ou previnir que durmam enquanto dirigem. Essa gama de aplicações ganha ainda mais relevância diante de estudos como o desenvolvido por Chang e Huang (2010), o qual atesta que as EFs compõem 55% da comunicação estabelecida entre os seres humanos, comprovando a sua importância nas relações interpessoais.

A fim de simplificar o escopo de análise no estudo das EFs, a maioria dos trabalhos nessa área seleciona para análise apenas as seis emoções (além da expressão neutra que geralmente é utilizada como referência inicial) classicamente consideradas universais por estudos realizados na Psicologia. Tais estudos consideram que este conjunto de emoções são suficientes para compreensão das relações e reações de emoções e sentimentos entre os seres humanos (WHISSELL, 1989). Essas seis emoções expressam: felicidade, surpresa, raiva, desgosto, medo e tristeza. Apesar desta tendência, um estudo recente (JACK; GARROD; SCHYNS, 2014) mostra que estas seis expressões poderiam ser reduzidas em quatro, categorizadas como: felicidade, tristeza, medo/surpresa e desgosto/raiva.

No contexto das Línguas de Sinais (LS), essas expressões são denominadas de Expressões Faciais Afetivas (EFA) (QUADROS; KARNOPP, 2004; LIDDELL, 2003), e assim como nas Línguas Orais (LO), têm o papel de demonstrar intenções e sentimentos durante o discurso, com a diferença que as LOs possuem a entonação da voz para colaborar nesse aspecto. Nas LSs, por serem línguas de modalidade visual-espacial e não contarem com o suporte sonoro da entonação, as EFAs ganham uma importância ainda maior, e outras EF, próprias desta modalidade de língua, colaboram para formar a estrutura gramatical da língua. Tais expressões são chamadas Expressões Faciais Gramaticais (EFG) e estão presentes nos níveis morfológico e sintático das LSs.

As EFGs em Nível Morfológico possuem o papel de adjetivação e modificam a semântica dos sinais das LSs enquanto estão sendo realizados. Elas podem atuar de maneira isolada, ou seja, por meio de alterações realizadas somente na EF durante a execução

do sinal (como no caso dos sinais de GORDO, GORDINHO e MUITO-GORDO); ou colaborando na adjetivação dos sinais por meio de alterações realizadas na EF junto com a modificação de outras unidades lexicais (como no caso da execução dos sinais de CASA, CASINHA e CASÃO, que além de sofrer modificação na EF, também sofre modificação nos movimentos realizados com mãos e braços durante a execução do sinal).

Com relação às EFGs no Nível da Sintaxe, a colaboração se dá na composição do sentido e da coesão do que está sendo sinalizado, expressando, por exemplo, negação, afirmação, interrogativas e expressões relativas. A ausência de EFGs durante o discurso nas LSs pode representar falta de fluência e/ou algum problema linguístico ocasionado durante a aquisição da LS como primeira língua ou ocasionado por problemas linguísticos desenvolvidos no nível neurolinguístico (LILLO-MARTIN; QUADROS, 2005).

Vale ressaltar que o hemisfério esquerdo do cérebro é responsável pelo processamento das EFGs, e esta é a mesma região que processa grande parte das funções linguísticas das línguas naturais, independente de sua modalidade oral ou visual-espacial. Já o hemisfério direito é responsável pelo processamento das EFAs, considerando suas manifestações nas LSs e nas LOs. Esta constatação sobre como as EFGs e as EFAs são processadas no cérebro ajuda a reafirmar a posição da LS como língua natural, pois os mesmos músculos são manipulados por regiões diferentes do cérebro por causa dos propósitos diferentes exigidos pelas funções nas quais as EFs são contextualizadas (QUADROS; KARNOPP, 2004).

Segundo Stokoe (1978), a simultaneidade dos constituintes das LSs (configurações de mãos, ponto de articulação, orientação/direção das mãos, movimento e expressões não-manuais) são marcas que as diferenciam das LO. Essas últimas possuem os fonemas como seus constituintes e eles são produzidos de maneira sequencial. Dentre os constituintes das LSs, as expressões não-manuais podem ser executadas por meio de EF, o que corrobora com o estabelecimento da importância delas dentro da fala e do discurso realizado por meio de LS.

As Expressões Não-Manuais (ENM), construídas por meio da posição e movimentos da cabeça, da posição e movimentos do corpo, do olhar e das EFs, estão presentes em LSs de outros países, como por exemplo no Japão (XU et al., 2000), Taiwan (SU; ZHAO; CHEN, 2001) e França (BRAFFORT, 1996). Elas ganham destaque no processo de reconhecimento automático das LSs, pois colaboram para retirada de ambiguidades entre sinais que possuem parâmetros semelhantes, como configuração de mãos e ponto de

articulação (GWETH; PLAHL; NEY, 2012), além de colaborarem na composição do sentido semântico das sentenças.

Os primeiros estudos que objetivam a interpretação das LSs eram focados na extração de características apenas dos sinais manuais, como pode ser observado em (KELLY et al., 2009a) que utiliza somente o movimento da cabeça para analisar os sinais manuais e não utiliza a informação das expressões faciais. Atualmente, muitos estudos passaram a considerar também as ENM. Alguns exemplos de estudos nesta linha são (AGRIS; KNORR; KRAISS, 2008), (CARIDAKIS; ASTERIADIS; KARPOUZIS, 2011), (KELLY et al., 2009b) e (KOSTAKIS; PAPAPETROU; HOLLMÉN, 2011). Com isso, o estudo da interpretação das LSs passou a ter, em muitos casos, um caráter multimodal devido à consideração da característica de simultaneidade dos constituintes das LSs. Esses e outros estudos são comentados em mais detalhes no Capítulo 3.

É nesse contexto de estudo das ENMs que este trabalho está inserido. De forma mais específica, este trabalho tem seu foco na reconhecimento de EFGs usadas na Língua Brasileira de Sinais (Libras). A fim de melhor apresentar o estudo desenvolvido neste trabalho, o restante desta introdução é dedicado a apresentar o objetivos delineados para o trabalho, uma definição resumida sobre como o reconhecimento das EFGs foi modelado, a metodologia escolhida para alcance desses objetivos, e, finalmente, como o restante do texto é organizado.

1.1 Objetivos

O principal objetivo desta dissertação de mestrado é desenvolver um conjunto de modelos de reconhecimento de padrões capazes de reconhecer as EFGs usadas no contexto da Língua Brasileira de Sinais (Libras), considerando-as em Nível Sintático. A fim de alcançar este objetivo, o problema de reconhecimento de padrões foi estudado sob dois diferentes aspectos:

1. Como um problema atemporal onde características de natureza espacial definem uma EFG tipicamente usada no contexto da Libras e podem ser identificadas em um *frame* de vídeo;

2. Como um problema temporal onde características de natureza espaço-temporal são necessárias para definir uma EFG tipicamente usada no contexto da Libras e devem ser identificadas a partir de um conjunto de *frames* de um vídeo.

Neste contexto, os seguintes objetivos específicos são delineados:

- Apresentar os conceitos fundamentais que formalizam o uso de EFGs dentro do contexto da Libras;
- Levantar o estado da arte na área de reconhecimento de EFs como elemento constituinte do discurso e do comportamento humano no contexto da comunicação e de relações interpessoais, i.e., considerando o reconhecimento de emoções expressadas por meio de EFs;
- Levantar o estado da arte na área de reconhecimento de EF como elemento constituinte da gramática e do discurso no contexto das LSs, i.e., considerando o reconhecimento de EFGs.
- Construir, e disponibilizar publicamente, um conjunto de dados referente ao uso de EGF durante a fala em Libras;
- Modelar o problema de reconhecimento de EFGs no contexto da Libras como um problema de classificação, considerando os aspectos espaço-temporais inerentes ao problema;
- Aplicar uma técnica de reconhecimento de padrões supervisionada no problema de reconhecimento de EFGs, avaliar os resultados sob a ótica de avaliação comumente aplicada na área de Aprendizado de Máquina e também sob a ótica de avaliação de especialistas da área de estudo de Línguas de Sinais.

1.2 Definição do problema

Nesta dissertação de mestrado, uma expressão facial $EF_i \in \{EF_1, EF_2, ...EF_n\}$ é a forma como um conjunto de pontos $P = \{p_1, p_2, ...p_n\}$, extraídos da face humana, estão dispostos no espaço tridimensional. Estes pontos possuem coordenadas (x, y, z), sendo o x uma coordenada em pixel no eixo horizontal, y uma coordenada em pixel do eixo vertical e z uma coordenada de profundidade dada em milímetros. A Figura 2 mostra a face neutra e um exemplo de EFG usada na Libras, considerando um *frame* real extraído de um vídeo onde uma EFG está sendo executada, e os respectivos pontos (x, y) extraídos da face.

(a) Face Neutra

(b) Pontos da Face Neutra

(c) Pontos da Face com a EF

Figura 1 – Exemplo de face neutra e da execução de uma EFG, e os respectivos pontos (x, y) extraídos da face.

A fim de definir o problema básico de classificação estudado neste trabalho, considere um vídeo como sendo uma sequência de *frames* $S = \{f_1, f_2, ..., f_n\}$ de tamanho n. Uma representação vetorial desse vídeo é usada como entrada para um modelo de classificação binária, cujo objetivo é classificar cada um dos *frames* como sendo referente ou não à execução de um EFG específica. Mais detalhes sobre a definição do problema e das componentes de sua solução são apresentados nos capítulos 5 e 6.

Vale ressaltar que não é objetivo desta dissertação de mestrado determinar a tradução semântica do que está sendo sinalizado, mas sim, identificar qual configuração a face assumiu durante a execução de uma fala em Libras. Assim, o reconhecimento aqui proposto assume um carater descritivo da LS. Como algumas EFGs possuem somente um significado semântico, naturalmente sua identificação corresponderá ao seu significado, como por exemplo, a EFG de negação, que pode implicar na modificação de um sinal ou constituir-se como uma negação genérica.

1.3 Metodologia

Para conhecer o estado da arte na área de reconhecimento das EF, tanto em relação ao estudo de emoções quanto no contexto do uso de EFGs no discurso em Libras, foram realizadas duas revisões sistemáticas (RS), seguindo a metodologia descrita em (KITCHENHAM, 2004). A RS é uma metodologia rigorosa que procura identificar o estado da arte de um determinado assunto por meio de coletas sistemáticas, combinações e avaliações críticas de trabalhos de uma determinada área (BIOLCHINI et al., 2005); um protocolo de pesquisa é estabelecido e direciona todo o processo de análise da literatura, além de permitir que a revisão seja passível de reprodução. Nessas revisões, foi possível

conhecer quais as técnicas e formas de análise comumente utilizadas nos contextos de reconhecimento de EFs e EFGs, além de conhecer os principais desafios e questões ainda em aberto relacionados ao tema.

As aquisições de dados nesta dissertação de mestrado foram realizadas com o uso do sensor Microsoft Kinect^{TM}[1] e do conjunto de funções disponibilizados na *Face Tracking SDK*[2]. Essa biblioteca possui um sistema de rastreamento de pontos específicos da face, ideal para a modelagem do problema estudado nesta dissertação de mestrado. Um dos motivos da escolha desta ferramenta é a possibilidade de aquisição da informação de profundidade dos pontos extraídos da face, de forma que um estudo abrangente pudesse ser feito em relação ao tipo de informação útil para o reconhecimento aqui pretendido. Além disso, são fatores importantes a condição de uso gratuito da biblioteca (*software*) e o baixo custo do equipamento de sensoriamento (*hardware*). Ainda em relação ao suporte de software para aquisição de dados, neste trabalho fez-se uso de *threads* paralelos do processador para compensar o custo computacional da captura *frame* a *frame*, aumentar o desempenho e maximizar o número de *frames* capturados por segundo.

Com relação à extração das características para criação de representações vetoriais para os vídeos capturados, além das próprias coordenadas dos pontos, foram usadas medidas de distâncias e ângulos entre os pontos. Diferentes representações vetorias, com combinações dessas características, foram analisadas a fim de que se pudesse indicar boas representações para resolução do problema de reconhecimento das EFGs.

Uma série de experimentos com a técnica de classificação *Multilayer Perceptron* (MLP) foram realizados, considerando: diferentes combinações das instâncias de dados existentes no conjunto de dados, diferentes representações vetoriais, uso de representação atemporal (um *frame*) e espaço-temporal (uma janela de *frames*) e diferentes combinações referentes aos parâmetros livres da MLP. A escolha da técnica MLP está baseada na sua boa capacidade de generalização, atestada por meio da literatura da área de Aprendizado de Máquina.

Para avaliação dos resultados obtidos com os classificadores, duas estratégias foram adotadas: (a) aferição de desempenho por meio de medidas classicamente usadas na avaliação de classificadores binários (HAN; KAMBER, 2006); (b) aferição de desempenho

[1] http://msdn.microsoft.com/en-us/library/hh855347.aspx
[2] http://msdn.microsoft.com/en-us/library/jj130970.aspx

por meio de análises derivadas do estilo de trabalho de especialistas em análise de gestos, a exemplo das análises realizadas por Madeo (2013).

1.4 Organização do documento

Este documento de dissertação está dividido em sete capítulos, considerando esta introdução. Os demais capítulos estão organizados da seguinte forma:

- O Capítulo 2 apresenta conceitos fundamentais sobre as LSs, além de contextualizar esta dissertação de mestrado e justificar sua importância para os estudos na área da surdez.
- O Capítulo 3 apresenta o levantamento bibliográfico referente ao reconhecimento de EFs, considerando o contexto de análise de emoções e o contexto de uso das EFs como EFGs no escopo das LSs. Esse levantamento representa o resultado consolidado de duas revisões sistemáticas.
- O Capítulo 4 apresenta um resumo sobre a teoria MLP, utilizada como estratégia de classificação aplicada neste trabalho, além de citar o algoritmo de treinamento com retropropagação de erros que é largamente utilizado neste contexto.
- O Capítulo 5 apresenta o conjunto de dados utilizado para os experimentos, além de detalhar a forma como os experimentos foram organizados.
- O Capítulo 6 apresenta os resultados dos experimentos, organizados por expressão facial, e também destaca quais características em comum foram identificadas em todos experimentos.
- O Capítulo 7 apresenta as considerações finais deste trabalho, com destaque para as principais contribuições do estudo e as possibilidades de extensão do mesmo.
- Por fim, nos Apêndices A e B são apresentados os protocolos usados nas revisões sistemáticas e os parâmetros envolvidos na condução da revisão e o Apêndice C traz os pontos da face fornecidos pelo Microsoft Kinect.

2 Expressões Faciais Gramaticais na Língua Brasileira de Sinais

As LSs, assim como as LOs, constituem-se como sistemas onde existem regras que definem padrões, geralmente estabelecidos naturalmente ao longo do tempo e do desenvolvimento da língua. Como discutido por Quadros e Karnopp (2004, p. 28), no caso específico de LS, esse sistema é composto por sinais arbitrários, e é caracterizado por uma estrutura definida (parâmetros), criativa e carregada de transmissão cultural. Tais características são inerentes a todas as línguas usadas no mundo, pois todas elas possuem semelhança em relação a seus elementos constituintes.

Alguns trabalhos sobre a Língua de Sinais Americana (ASL – do inglês *American Sign Language*) (FRISHBERG, 1975; JR, 1976), mostram a variação da LS com o tempo, devido a ocorrência da transmissão cultural e também a adaptação causada pelas novas experiências proporcionadas, por exemplo, pelas mudanças tecnológicas. Esses estudos mostram também que, ao contrário do que muitos pensam, a representação das formas visuais (mímicas ou iconicidade) não são relevantes na formação de novos vocabulários nas LSs.

Dentro das regras e práticas presentes na LS existem aquelas que caracterizam os aspectos morfológicos, gramaticais e semânticos da língua. Esses aspectos, por sua vez, permitem a construção do discurso da fala. O discurso, segundo Dubois (2001, p. 192), "é a linguagem posta em ação, sendo considerado unidade igual ou superior a frase, constituído por uma sequência de começo, meio e fim. Na concepção da Linguística Moderna, o termo discurso designa todo enunciado superior à frase, considerado do ponto de vista das regras de encadeamento das sequências de frases."

Considerando este contexto, a morfologia estuda aspectos da estrutura e classificação dos itens lexicais de uma língua (palavras nas LOs e sinais nas LSs), bem como suas classificações, analisando o item isoladamente, sem se preocupar com o contexto ou frase no qual ele é utilizado. A sintaxe analisa como esses itens são usados na formação de uma frase e como as frases compõem o discurso. Além disso, faz parte também da tarefa do estudo sintático, entender a relação lógica existente entre as frases no discurso. Já a semântica tem o objetivo de estabelecer, ou interpretar, o significado das estruturas sintáticas na linguagem.

As EFs fazem parte da comunicação humana, independentemente se o sistema linguístico usado é oral ou gestual. Principalmente por meio das EFs, as pessoas expressam emoções ou intenções, e portanto, modificam o discurso expresso em sua fala. Especificamente em uma LS, tanto a morfologia, quanto a sintaxe e a semântica são definidas pelo uso conjunto de diferentes elementos gestuais e entre eles estão as EFs – ou seja, além de ser naturalmente usada para expressar emoções ou intenções, as EFs nas LSs assumem um papel de maior destaque, sendo essenciais para dar sentido ao que é dito (FERREIRA-BRITO, 1995).

Este capítulo tem o objetivo de explicar a importância das EFs no contexto das LSs, tomando como base a Língua Brasileira de Sinais (Libras). Ele está organizado da seguinte forma: os conceitos fundamentais sobre LSs são apresentados na Seção 2.1; a Libras é brevemente apresentada na Seção 2.2; o papel das EFs no que diz respeito ao estudo gramatical da Libras, bem como seus efeitos no discurso, são discutidos da Seção 2.3. É importante, para este trabalho, considerar o contexto de uma LS em específico, uma vez que embora o estudo realizado em uma língua possa ser localizado em outra, algumas estruturas morfológicas, sintáticas e semânticas são particulares a cada uma.

2.1 Língua de Sinais

As primeiras pesquisas sobre as LSs foram realizadas em meados de 1960, por William C. Stokoe Jr.. Em sua obra (STOKOE, 1978), Stokoe identificou três parâmetros básicos da ASL, e defendeu que eles estão presentes em todas as LSs. São eles:

1. Configuração das mãos: formato que a mão assume durante a execução dos sinais;
2. Ponto de articulação: local onde o sinal é realizado no espaço tridimensional, "ancorado" ao corpo ou realizado no espaço neutro em frente do interlocutor;
3. Movimento: movimento realizado pelas mãos, no espaço tridimensional, durante a sinalização.

Tais parâmetros não são necessariamente executados em todos os sinais e também não precisam ocorrer sempre ao mesmo tempo.

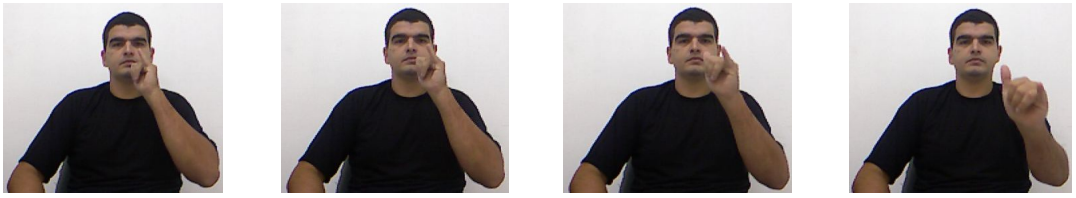
Mais tarde, Battison (1974), Ferreira-Brito (apud QUADROS; KARNOPP, 2004) e Aarons (1994) argumentaram que o estudo das LSs seria mais complexo. Battison introduziu a importância da observação de aspectos de orientação e direção nos parâmetros básicos,

defendendo que estes aspectos comporiam um novo parâmetro na composição dos sinais. E, a partir do estudo sobre traços[1] não-manuais, Ferreira-Brito e Arons introduziram o quinto parâmetro básico para o sistema: as expressões não-manuais. Assim, dois parâmetros foram adicionados no conjunto de parâmetros básicos para a composição do sistema gestual:

4. Orientação e direção: orientação que a palma da mão assume durante a produção do sinal e a direção que esse sinal é executado;
5. Expressões não-manuais: posição e movimentos da cabeça, posição e movimentos do corpo, olhar e EFs.

Segundo Stokoe (1978), a diferença fundamental entre as LSs e as LOs está relacionada com a estrutura simultânea de organização dos elementos que compõem as unidades lexicais da língua – os sinais. Segundo o mesmo autor, nesse sentido, os sinais podem ser vistos como um conjunto de traços mínimos não-holísticos que não possuem significado isoladamente.

Como um exemplo de um sinal, sob a perspectiva de Stokoe, observe a Figura **??**. A configuração de mão está em **Y**, a palma da mão está virada para a esquerda, o ponto de articulação está saindo da boca com um movimento e direção responsáveis por dizer quem é o ativo e o passivo da frase. O significado do que está sendo dito neste exemplo é EU DISSE VOCÊ. Essa configuração de mão por si só não possui nenhum significado na Libras, então é necessário que a mesma seja executada com outros parâmetros para que tenha significado.



(a) DIZER(i) (b) DIZER(ii) (c) DIZER(iii) (d) DIZER(iv)

Figura 2 – Sinal DIZER sendo executado em partes. Note que ao incluir movimento e direção na composição do sinal, os sujeitos EU e VOCÊ se tornam parte do significado do que está sendo dito.

[1] Aspecto mínimo de um sinal, por exemplo, a configuração de mão.

2.2 Língua Brasileira de Sinais

A Língua Brasileira de Sinais (Libras) é resultado da influência da Língua de Sinais Francesa na cultura brasileira. Essa LS foi trazida para o contexto brasileiro pelo conde francês Ernest Huet, em 1856 (com a vinda de membros da corte portuguesa para o Brasil). Naquela época, várias LSs já eram utilizadas em diferentes regiões do Brasil, e estas também influenciaram a formação da Libras.

A Libras ganhou força no país em 1857, quando foi fundado o Instituto dos Surdos-Mudos do Rio de Janeiro, hoje Instituto Nacional de Educação de Surdos (INES), de onde saíram, e ainda saem, agentes divulgadores da Libras. Hoje, a Libras é reconhecida como a língua oficial para comunicação e expressão da comunidade surda do Brasil pela Lei nº 10.436, de 24 de abril de 2002.

Assim como qualquer LS, a Libras possui um sistema de sinais gestuais bastante rico e complexo. Ainda hoje, os pesquisadores da área de estudo da Libras não estabeleceram um consenso sobre um conjunto finito de instâncias dos elementos básicos constituintes dos sinais usados na língua. Porém, visto que a presente proposta de trabalho objetiva fazer uma análise automatizada de um desses elementos, faz-se interessante fornecer uma ideia sobre a dimensão da complexidade assumida pelos estudiosos da Libras.

De acordo com Ferreira-Brito (1990), na Libras há 46 configurações de mão, 6 tipos de orientações de mão, em torno de 40 locações no corpo, 16 locações no espaço neutro[2], 23 expressões não-manuais, uma lista de 35 possíveis movimentos internos da mão e, com relação ao movimento, Ferreira-Brito identifica 28 especificações (retilíneo, circular, curvo, etc.), 17 direcionalidades, 5 maneiras (qualidade, tensão e velocidade, que pode ser rápido, mais tenso ou mais frouxo) e 2 tipos de frequência (simples ou com repetição).

Entretanto, como afirmado por Amaral (2012), ainda não existe um levantamento exaustivo sobre todos os possíveis estados que os elementos constituintes de uma LS podem assumir no contexto da Libras. Assim, é sabido que não existe um consenso sobre esta questão. Ainda segundo a mesma autora, "o banco de dados de sinais da Libras desenvolvido por Xavier (apud AMARAL, 2012) permite-nos ter uma boa perspectiva sobre quais e quantas são as configurações de mão presentes na Libras".

[2] Locações 'ancoradas' no corpo ou no espaço neutro fazem parte do conjunto de pontos de articulação, um dos elementos constituintes de uma LS.

Cada um desses elementos podem ser explorados de forma mais aprofundada para que se tenha a real compreensão sobre todos os aspectos que envolvem o estudo da língua e o entendimento de suas especificidades. Porém, o presente trabalho tem o objetivo de explorar apenas um deles - a Expressão Facial. Assim, o restante deste capítulo trata de aspectos referentes ao uso da EF na composição de sinais e frases considerando o contexto da Libras, provendo as informações necessárias para compreensão do problema e das soluções tratadas nesta dissertação.

2.3 Expressões Faciais Gramaticais da Libras

Nos estudos clássicos da Psicologia das Relações Humanas, Ekman (1978) cita a existência de seis emoções, perceptíveis por meio da EF, que podem ser consideradas universais na compreensão das reações e sentimentos que permeiam as relações entre seres humanos: felicidade, surpresa, raiva, desgosto, medo e tristeza, além da expressão neutra que geralmente é utilizada como uma referência inicial para a análise das demais. Para (JACK; GARROD; SCHYNS, 2014) estas expressões são reduzidas em quatro, sendo elas: felicidade, tristeza, medo/surpresa e desgosto/raiva.

No contexto das LSs, a interpretação das EFs vai além da análise das emoções. De fato, considera-se que as expressões que representam emoções são chamadas de Expressões Faciais Afetivas (EFA) e embora elas possam ocorrer durante o discurso, elas estão ligadas à ele no mesmo nível em que estão quando acompanham a fala oral. Para além deste significado ou função, as EFs na LS podem ser "gramaticais".

Essas expressões, chamadas Expressões Faciais Gramaticais (EFG) estão dentro do contexto de Expressões Não-Manuais (ENM), sendo caracterizadas por: posição e movimentos da cabeça, posição e movimentos do corpo, olhar e EF. As EFGs, segundo Quadros e Karnopp (2004) e Arroteia (2005), estão relacionadas as estruturas específicas da LS, tanto no nível da morfologia quanto no nível da sintaxe, e são obrigatórias em determinados contextos. Elas podem, então, ser usadas para modificar sinais, impondo o que se chama de alteração em alguns dos seus traços mínimos que levam à alteração do sentido da frase que está sendo dita.

As EFGs que modificam um dos parâmetros básicos (ENM) de composição do sistema gestual estão no centro de interesse de estudo deste trabalho.

É interessante notar que estudiosos de LSs têm a prática de analisar os elementos básicos da língua de forma detalhada. Ferreira-Brito (1990), por exemplo, desenvolveu um estudo sobre quais são as possibilidades de movimentos para formação de ENM, e como cada um desses movimentos é importante para a formação da expressão no uso da língua. Na Tabela 1 são destacadas as possibilidades discutidas em Ferreira-Brito (apud QUADROS; KARNOPP, 2004) com relação a cabeça e face.

Tabela 1 – Possibilidades de movimentos importantes para a construção da EF na fala em LS. Adaptado de Ferreira-Brito (apud QUADROS; KARNOPP, 2004).

Rosto	
Parte Superior	*Parte Inferior*
sobrancelhas franzidas	bochecha(s) inflada(s)
sobrancelhas levantadas	bochecha(s) contraída(s)
olhos arregalados	lábio(s) contraído(s)
lance de olhos	lábio(s) projetado(s)
	franzir do nariz
	movimento da língua
Cabeça	
movimento para frente e para trás	inclinação para frente e para trás
movimento para os lados	inclinação para os lados
Rosto e Cabeça (combinados)	
cabeça projetada para a frente, olhos levemente cerrados e sobrancelhas franzidas	
cabeça projetada para trás e olhos arregalados	

Segundo Quadros e Karnopp (2004), as expressões faciais gramaticais têm a função de executar marcações não-manuais e podem ser divididas em dois níveis:

- nível morfológico: marcações não-manuais que acompanham um adjetivo ou substantivo, determinando o grau de intensidade quando associadas ao adjetivo e o grau de tamanho quando associadas ao substantivo, e permitindo a construção de superlativos e comparativos de superioridade e inferioridade;
- nível da sintaxe: marcações não-manuais responsáveis por construir sentenças negativas, interrogativas, afirmativas, condicionais, relativas, com tópicos e com foco.

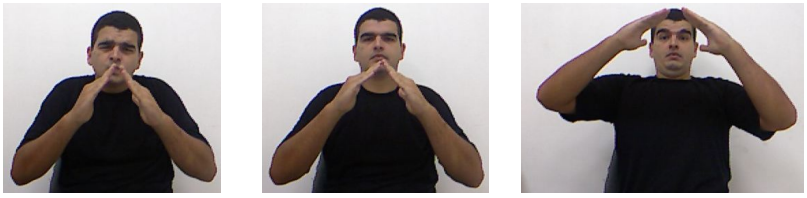
2.3.1 Nível Morfológico

Seguindo o exposto em (QUADROS; KARNOPP, 2004), no nível morforlógico, as EFGs têm a função de impor grau de adjetivação. O uso e efeito do uso dessas expressões podem acontecer de maneira isolada, ou seja, apenas a modificação do traço do elemento básico "expressão facial" é suficiente para indicar alguma alteração na intensidade que se quer impor ao objeto de adjetivação; ou acompanhado de modificações em outros elementos básicos (na configuração da mão ou no movimento, por exemplo) da construção do sistema gestual.

Como exemplo do primeiro caso, considere o adjetivo GORDO. A intensidade imposta ao adjetivo, resultando nos superlativos GORDINHO ou GORDÃO, é manifestada nas variações na EF (veja Figura 3). Para o segundo caso, considere o substantivo CASA. As variações na EF apoiam a modificação dos traços do sinal em relação a outros parâmetros, como a configuração da mão, para construção dos superlativos CASINHA e CASARÃO (veja Figura 4).

(a) GORDINHO (b) GORDO (c) GORDÃO

Figura 3 – EF impondo intensidade a um adjetivo.



(a) CASINHA (b) CASA (c) CASARÃO

Figura 4 – EF apoiando a modificação dos traços do sinal para um substantativo.

Observa-se o uso das EFs em nível morfológico nas Figuras 3 e 4, sendo interessante notar que as expressões no diminutivo para GORDINHO e CASINHA são as mesmas, caracterizadas pela contração dos lábios, olhos e sobrancelhas, enquanto as expressões no aumentativo possuem configurações opostas à contração, com as bochechas infladas para GORDÃO e olhos arregalados para CASARÃO. Vale ressaltar que o "nível" de intensidade

das expressões é particular ao sinalizante, assim como a ênfase dada ao que se deseja transmitir em sua fala.

2.3.2 Nível Sintático

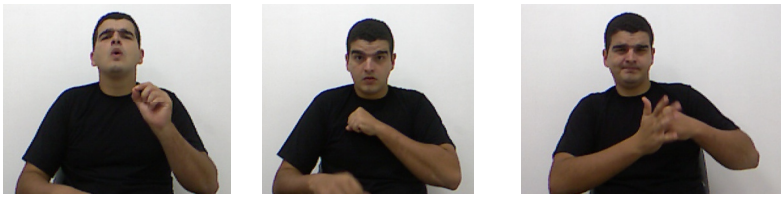
No nível sintático, as EFGs são responsáveis por construir frases interrogativas e frases relativas, determinar polaridade (frases afirmativas e negativas) e condicionais, elaborar construções com tópico e com foco. É interessante notar que a sinalização dos mesmos sinais que compõem uma construção dentre as citadas, se repetida sem a execução da EF, tornaria a frase agramatical.

A fim de explicar em mais detalhes o uso das EFGs nos diferentes contextos de construção de frases, segue uma discussão ilustrada baseada nos exemplos apresentados por Quadros e Karnopp (2004) e Ferreira-Brito (1990):

1. Interrogativa: Frases formuladas com a intenção de obter alguma informação desconhecida. Há quatro tipos de EFs interrogativas responsáveis por esse contexto. A Figura 5 ilustra os três primeiros tipos, sendo que o último incorpora outras EFs que também estão ilustradas nesta seção.

    - Interrogativas (qu): por meio do uso de expressões interrogativas do tipo QUEM, QUE, QUANDO, POR QUE, COMO, ONDE. Esta expressão é caracterizada por uma pequena elevação da cabeça acompanhada do franzir da testa. Por exemplo: <QUE JOÃO PAGAR>(qu)[3]
    - Interrogativas (s/n): formula questões que esperam como resposta um SIM ou um NÃO. Como por exemplo: <JOÃO COMPRAR CARRO>(s/n). Percebe-se visualmente que há um abaixamento da cabeça e uma elevação das sobrancelhas.
    - Interrogativas (dúvida): expressam algum tipo de desconfiança. Por exemplo: <JOÃO BANHEIRO TRANCADO>(dúvida). Nesta expressão os lábios ficam comprimidos, os olhos mais fechados, testa franzida e há uma leve inclinação dos ombros para o lado ou para trás.
    - QUE ou QUEM aparecendo em sentenças subordinadas sem a EF interrogativa, utilizando a marcação própria da frase. Por exemplo: <EU SEI QUEM ROUBOU> (afirmativa).

[3] Notação para indicar que a EF de Interrogação (qu) foi utilizada em toda frase, sendo que <> delimita o período de execução da expressão.

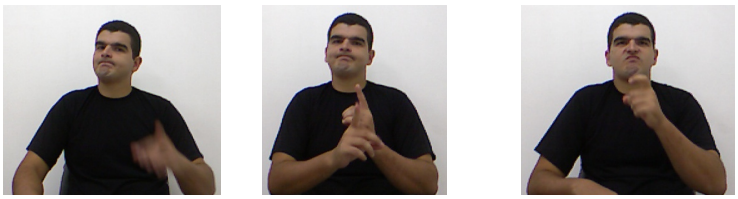


(a) qu (b) s/n (c) dúvida

Figura 5 – EFs usadas em frases interrogativas.

2. Negativa: Normalmente possuem um elemento negativo explícito, como NÃO, NADA, NUNCA que, na LS, podem estar incorporados aos sinais ou expressos apenas por meio da marcação não-manual (ARROTEIA, 2005). A Figura 6 traz três exemplos de EFs usadas em frases negativas.

   Essas EFs podem ser descritas visualmente pelo movimento horizontal da cabeça, ou pelo movimento do contorno da boca (abaixamento dos cantos) sempre associados ao abaixamento das sobrancelhas e abaixamento da cabeça.



(a) qu (b) s/n (c) dúvida

Figura 6 – EFs usadas em frases negativas.

3. Afirmativa: frases que expressam ideias ou ações afirmativas usando, por exemplo, <EU VOU SHOPPING>(afirmativa). Como pode ser visto na Figura 7, essas expressões são caracterizadas pelo movimento vertical da cabeça para cima e para baixo.

(a) qu (b) s/n (c) dúvida

Figura 7 – EFs usadas em frases afirmativas.

4. Condicional: frases que estabelecem uma condição para realizar alguma coisa, por exemplo, <SE CHOVER>(condição) <EU NÃO VOU FESTA>(negativa). A Figura 8 ilustra como uma EF é colocada nesse contexto, caracterizada pelo abaixa-

mento da cabeça e suspensão das sobrancelhas durante a execução da condição, seguida imediatamente por uma outra EF, podendo ser negativa ou afirmativa.

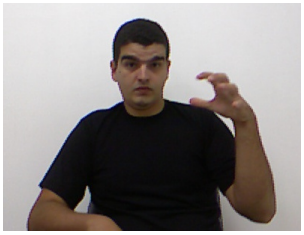
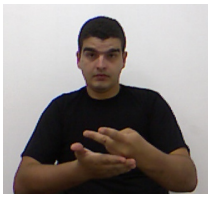
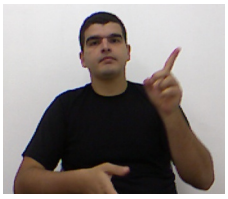


Figura 8 – EF usada em construções com condicionais.

5. Relativa: trata-se de uma inserção dentro da frase para explicar, acrescentar informações ou encaixar outra questão relativa ao que está sendo dito (QUADROS; KARNOPP, 2004). A Figura 9 ilustra essa EF, caracterizada por uma quebra de expressão entre a informação inserida (elevação da sobrancelha) e o restante da frase, como por exemplo: <MENINA CAIU BICICLETA>(relativa) ELA LÁ HOSPITAL.



(a) qu



(b) s/n

Figura 9 – EFs usadas em frases relativas.

6. Tópicos: forma diferente de organizar o discurso, por exemplo, <FRUTAS> (tópico) EU GOSTO BANANA, i.e., o tópico é FRUTAS, nesse contexto "eu gosto de banana". Essas expressões podem ser caracterizadas de três maneiras diferentes:

   a) elevação das sobrancelhas, cabeça inclinada para baixo e para o lado e olhos arregalados;
   b) movimento da cabeça para baixo e para frente, olhos arregalados seguido de amplo movimento da cabeça para trás e para o lado;
   c) cabeça para frente, inclinada levemente para cima ou para baixo, boca aberta com a parte superior elevada, sobrancelha elevada e olhos arregalados.

7. Foco: frases que introduzem no discurso uma informação nova que pode: (a) estabelecer contraste; (b) informar algo adicional; (c) enfatizar alguma coisa. Por exemplo, se alguém diz que a MARIA COMPROU O CARRO e essa informação

Figura 10 – EF utilizada em frase com Tópico (b).

está equivocada, pode-se ter a seguinte frase logo depois: NÃO <PAULO> (foco) COMPROU CARRO. (QUADROS; KARNOPP, 2004). Esta EF é caracterizada pela mesma EF utilizada em tópico (a), Figura 11 mostra como ela é utilizada.

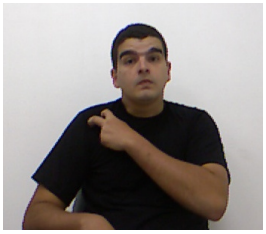


Figura 11 – EF utilizada em frase com Foco.

2.3.3 Considerações Finais

Este capítulo, além de trazer uma breve explanação sobre a Libras, também forneceu informações com o intuito de contextualizar o problema tratado nesta dissertação. No decorrer do capítulo, foram mostradas quais são as EFGs, e suas características, que compõem o conjunto de dados sob análise neste trabalho. Estas expressões estão presente no discurso da Libras e são utilizadas naturalmente por pessoas fluentes nessa LS.

3 Reconhecimento de Expressões Faciais

O problema estudado neste trabalho de dissertação diz respeito, especificamente, ao reconhecimento de EFGs no contexto da Libras. Este problema exige um grau de generalização que permita a análise de uma expressão facial de acordo com o que ela representa em um contexto, independentemente do indivíduo que a realiza. Reconhecer o significado de uma expressão facial envolve o estudo de características que a representam, seja ela uma expressão contextualizada em uma análise referente às LSs, LOs ou, de forma mais abrangente, ao estudo do comportamento humano em geral.

Assim, como parte do estudo desenvolvido para proposição de uma solução para o problema objetivado neste trabalho, foi realizado um levantamento bibliográfico sobre iniciativas que tiveram como objetivo a análise de emoções expressas por meio de EFs humanas, aqui tratadas como EFAs, e também estudos que objetivaram o estudo específico de reconhecimento de EFGs.

Este capítulo tem como objetivo oferecer um panorama geral do que vem sendo realizado na área de pesquisa correlata à análise de EFAs e EFGs, ilustrando o contexto de uso das EFs como forma de expressar emoções, e também considerando o contexto de inserção de EF na construção da fala em LS.

Para melhor organizar as informações resultantes deste levantamento, o presente capítulo está estruturado de forma a apresentar em um primeiro momento (Seção 3.1), as iniciativas voltadas para a análise das EFs humanas e reconhecer emoções, excluindo iniciativas voltadas para a combinação destas com outras formas de expressão corporal. A segunda parte do capítulo (Seção 3.2) traz informações de estudos que analisam as EFs dentro do escopo das LS, com destaque para o fato que muitos trabalhos utilizam os resultados obtidos no reconhecimento das EFs durante o discurso das LSscomo forma de melhorar os resultados do reconhecimento automático dos sinais.

Cada uma das duas partes deste capítulo, segue a seguinte organização de apresentação das informações referentes aos estudos analisados: apresentação do escopo das pesquisas realizadas pelos autores dos estudos verificados considerando as EFs utilizadas e as aplicações construídas a partir dos resultados das análises das expressões (Seções 3.1.1 e 3.2.1); apresentação das estratégias de aquisição de dados e técnicas de pré-processamento de dados comumente aplicadas (Seções 3.1.2 e 3.2.2); breve descrição de algumas bases de dados específicas para disponibilização de massa de dados referente a EFs (Seções 3.1.3

e 3.2.3); listagem das técnicas computacionais aplicadas na construção dos modelos que analisam as EFs e reconhecem as emoções, bem como a extração de características e representação de dados para uso nestas técnicas (Seções 3.1.4 e 3.2.4) e, finalmente, comentários sobre as metodologias de avaliação de desempenho dos modelos construídos em cada um dos artigos sob análise (Seções 3.1.5 e 3.2.5).

3.1 Expressões Faciais Afetivas

Segundo Ekman (1978), por meio de uma EF é possível perceber a execução de seis emoções (felicidade, surpresa, raiva, desgosto, medo e tristeza), as quais podem ser consideradas universais dentro do estudo e análises das relações humanas. Embora a maioria dos estudos que compõem este levantamento de literatura (25 estudos dos 35 analisados – 71%) tenham sido desenvolvidos sobre essas seis emoções, existem autores que defendem um número menor de expressões universais (JACK; GARROD; SCHYNS, 2014), e também aqueles que inserem em seus estudos, outras variações de expressões. Além disso, uma expressão considerada "neutra" é, geralmente, adicionada ao conjunto de expressões analisadas de forma a representar uma referência inicial para as demais análises.

3.1.1 Escopo dos Estudos

Dentro do conjunto de estudos analisados, há autores que optam por fazer análise de um conjunto pequeno de EFs, e há aqueles que elaboram um conjunto maior, a depender do contexto de aplicação no qual os autores motivam o seu estudo, e do tipo de emoção ou comportamento humano a ser analisado. Dentre os estudos nos quais os autores optam por reduzir o conjunto de EFs analisadas estão: (HUANG; LIN, 2008) no qual os autores analisaram somente as EFs relacionadas às emoções de surpresa, felicidade e raiva, e a expressão neutra; (TEWS et al., 2011) no qual os autores trabalharam com as expressões para felicidade, raiva e neutra; (SONG et al., 2008) onde os autores analisaram as expressões neutra, alegria, raiva, surpresa, tristeza e medo, por meio de uma análise multimodal, ou seja, além de imagens, informação de áudio foi incorporada à entrada dos modelos de análise; (SIDDIQUI; LIAO; MEDIONI, 2009) no qual os autores deixaram somente a EF de raiva fora de seu escopo; e (CHO; PARK, 2011) no qual somente as expressões de felicidade, surpresa e tristeza fazem parte do escopo analisado.

Alguns estudos, por outro lado, analisam um conjunto maior de EFs ou analisam EFs fora do escopo específico das emoções classicamente definidas. Neste conjunto de estudos estão: (PETRIDIS; PANTIC, 2011) no qual os autores estão interessados em diferenciar risadas de falas durante o discurso; (POPA; ROTHKRANTZ; WIGGERS, 2010) onde 21 EFs para analisar comportamento de consumidores em um supermercado são estudadas; (BOUCENNA; GAUSSIER; HAFEMEISTER, 2011) e, finalmente, (HOEY; LITTLE, 2007), nos quais os autores utilizaram EFs sem rotulação (executando uma análise de agrupamento) para analisar o comportamento das pessoas levando em consideração o contexto dos dados.

As aplicações que a análise de EF e reconhecimento de emoções possibilitam são discutidas em alguns trabalhos. Os autores em (VALENTI; JAIMES; SEBE, 2008) criaram uma interface gráfica para mostrar o reconhecimento automático de EF e, em (VALENTI; JAIMES; SEBE, 2010), foi proposta a criação de uma ferramenta que reproduz um som de acordo com a EF reconhecida. Já os autores de (JOHO et al., 2009) e (ZHAO et al., 2011) trabalharam com a classificação de gênero de vídeos por meio das EFs executadas por pessoas enquanto assistiam aos vídeos. A construção de uma casa inteligente, onde decisões são tomadas com base na EF do morador é proposta em (YU; YOU; TSAI, 2012).

3.1.2 Natureza dos Dados e Pré Processamento

Durante a análise dos artigos foi possível encontrar um conjunto comum de informações sobre os dados utilizados nas experimentações e aplicações desenvolvidas pelos pesquisadores. Essas informações, relacionadas na Tabela 2, dizem respeito a:

1. **Natureza:** informação sobre a proveniência dos dados (imagens (I) ou vídeos (V)).
2. **Abordagem:** informação sobre a forma como as informações foram trabalhadas (*pixels* (P)) ou informações sobre a geometria (G) do rosto humano.
3. **Dimensões:** dimensionalidade do espaço de extração das informações (duas dimensões - 2D ou três dimensões - 3D).
4. **Proveniência:** informação sobre a proveniência do conjunto de dados utilizados (primários (P) – adquiridos pelos próprios autores; secundários (S) – provenientes de outras fontes, como bases de dados públicas, por exemplo).
5. **Pré-processamento:** informações sobre se os dados são provenientes de estúdios (L) ou são usados em ambientes externos (E).

Tabela 2 – Informações sobre os dados utilizados nas experimentações e aplicações dos estudos referentes as EFAs.

Estudo	1		2		3		4		5	
	I	V	P	G	2D	3D	P	S	L	E
(HOEY; LITTLE, 2007)	x	-	x	-	x	-	-	x	x	-
(BUENAPOSADA; MUÑOZ; BAUMELA, 2008)	x	-	x	-	x	-	-	x	x	-
(HUANG; LIN, 2008)	x	-	x	x	x	-	x	-	-	x
(SONG et al., 2008)	x	-	-	x	x	-	-	x	x	-
(VALENTI; JAIMES; SEBE, 2008)	x	-	-	x	-	x	x	-	-	x
(XIANG; LEUNG; CHO, 2008)	x	-	x	-	x	-	-	x	x	-
(YANG; CHIANG, 2008)	-	x	-	x	x	-	-	x	x	-
(ZHOU; LIANG; ZHU, 2008)	-	x	x	-	x	-	-	x	x	-
(ZHI; RUAN, 2008)	x	x	x	-	x	-	-	x	x	-
(JOHO et al., 2009)	x	-	-	x	-	x	x	-	-	x
(SIDDIQUI; LIAO; MEDIONI, 2009)	x	-	x	-	-	x	x	-	-	x
(CHANG; HUANG, 2010)	-	x	x	x	x	-	-	x	x	-
(LAJEVARDI; HUSSAIN, 2010)	-	x	x	-	-	x	-	x	x	-
(LI; RUAN; LI, 2010)	x	x	x	-	x	-	-	x	x	-
(POPA; ROTHKRANTZ; WIGGERS, 2010)	x	-	x	x	x	-	x	x	x	x
(RUDOVIC; PATRAS; PANTIC, 2010)	-	x	-	x	x	-	-	x	x	-
(WANG et al., 2010)	x	x	-	x	x	-	x	-	-	x
(VALENTI; JAIMES; SEBE, 2010)	x	-	-	x	-	x	x	-	-	x
(BOUCENNA; GAUSSIER; HAFEMEISTER, 2011)	x	-	-	x	x	-	-	x	-	x
(CHO; PARK, 2011)	-	x	x	-	x	-	-	x	x	-
(GUO; RUAN, 2011)	x	x	x	-	x	-	-	x	x	-
(LEMAIRE et al., 2011)	-	x	-	x	-	x	-	x	x	-
(LIU; RUAN; WANG, 2011)	x	x	x	x	x	-	-	x	x	-
(PETRIDIS; PANTIC, 2011)	x	-	-	x	x	-	-	x	-	x
(TEWS et al., 2011)	x	-	-	x	x	-	x	-	x	-
(WANG et al., 2011)	-	x	x	-	x	-	-	x	x	-
(WU; SHEN; FU, 2011)	x	-	x	-	x	-	-	x	x	-
(YONG; SUDIRMAN; CHEW, 2011)	x	-	x	x	-	x	x	-	-	x
(ZHANG; GENG, 2011)	x	x	x	-	x	-	-	x	x	-
(ZHANG; TJONDRONEGORO; CHANDRAN, 2011)	-	x	x	-	x	-	-	x	-	x
(ZHAO et al., 2011)	x	-	x	-	x	-	x	x	-	x
(SARVADEVABHATLA et al., 2011)	x	-	x	-	x	-	-	x	x	-
(CHEN et al., 2012)	-	x	x	x	x	-	-	x	x	-
(DAHMANE; MEUNIER, 2012)	-	x	x	-	x	-	-	x	x	-
(YU; YOU; TSAI, 2012)	x	-	x	x	x	-	x	-	-	x

Dentre os artigos analisados destacam-se alguns que apresentam peculiaridades referentes ao dado sob exploração. Em (LAJEVARDI; HUSSAIN, 2010), a identificação das EFs é realizada a partir de imagens de baixa resolução; imagens com oclusões são analisadas em (ZHANG; TJONDRONEGORO; CHANDRAN, 2011); e, finalmente, o estudo de Rudovic, Patras e Pantic (2010) é direcionado à exploração de problemas com reconhecimento de EFs a partir de imagens com rostos não frontais.

Muitos autores não explicitam quais técnicas foram utilizadas na etapa de pré-processamento dos dados, mas percebe-se que a normalização das imagens é o procedimento mais comum. Nesse procedimento, os dados sofrem transformações para que possuam uma padronização de tamanho, intensidade luminosa, cor (escala cinza ou colorida) e resolução.

Dentre os trabalhos analisados, alguns esclarecem como foi realizada a segmentação da face na imagem. Por exemplo: Zhang e Geng (2011) aplicam a extração de quatro regiões da face (boca, nariz, olhos e região entre os olhos); Liu, Ruan e Wang (2011) propõem um novo tipo de ranqueamento ortogonal baseado nas variações intraclasses e interclasses; Chen et al. (2012) apresentam um sistema híbrido de segmentação baseado em formas geométricas e informação de textura das imagens; Lemaire et al. (2011) trabalham com dez marcações na face para sua análise; Guo e Ruan (2011) propõem um descritor de EF que permite analisar a imagem e encontrar a face; e Li, Ruan e Li (2010) apresentam um modelo baseado em Transformadas de *Wavelet*.

Li, Ruan e Li (2010) usaram procedimentos manuais para segmentar a face da imagem de entrada, e outros autores utilizaram bibliotecas com algoritmos de detecção automática de face, como Dahmane e Meunier (2012), Chen et al. (2012), Popa, Rothkrantz e Wiggers (2010) que utilizaram o algoritmo de *Viola-Jones* (VIOLA; JONES, 2001) e Wang et al. (2010) que utilizaram as funções disponibilizadas na *OpenCV*[1]. Já Tews et al. (2011) utilizaram marcadores colocados na face imageada com objetivo de facilitar o rastreaemento de pontos de interesse.

Outra atividade comum nos trabalhos analisados é a determinação de um ponto de referência, a linha dos olhos por exemplo, para padrozinar o alinhamento de todos os dados de entrada. Vale lembrar que alguns trabalhos que utilizaram dados secundários, provenientes de bases de dados publicadas, já receberam os dados de entrada tratados e não houve a necessidade de pré-processamento.

[1] `http://opencv.org/`

3.1.3 Bases de Dados

Duas bases de dados públicas foram amplamente utilizadas nos trabalhos analisados e esta subseção as descreve resumidamente. Em alguns trabalhos, ambas bases foram usadas, sendo que geralmente uma base foi utilizada para treinamento de modelos de reconhecimento e a outra para prover dados para teste e validação dos modelos de reconhecimento. As bases de dados são:

- **Jaffe**: Esta base de dados foi utilizada em dez dos trabalhos analisados (29%). Ela possui 213 imagens de sete EFs diferentes (seis expressões universais e uma expressão neutra) de dez modelos femininos japoneses (LYONS et al., 1998). A base de dados encontra-se disponível em `www.kasrl.org/jaffe.html`.
- **Cohn-Kanade**: A Cohn-Kanade (CK) é uma base de dados de EFs criada para auxiliar pesquisas em análise e síntese automática de faces, e também para pesquisa na área de análise perceptual. Ela foi utilizada em 11 dos artigos analisados (31%) neste levantamento. CK está disponível em duas versões, sendo que somente Chen et al. (2012) utilizou a versão mais nova. A primeira versão (original) inclui 486 sequências de 97 poses. Cada sequência começa por uma expressão neutra e prossegue para uma expressão máxima das seis expressões consideradas universais (KANADE; COHN; TIAN, 2000). A segunda versão, conhecida como CK+, inclui as expressões já existentes na CK original e um acréscimo de 22% nas sequências e 27% no número de indivíduos. Rótulos foram acrescentadas nas expressões e, além disso, a CK+ fornece protocolos e resultados obtidos em experimentos de classificação de expressões. Uma descrição completa da base de dados pode ser encontrada em (LUCEY et al., 2010) ou em `www.pitt.edu/~jeffcohn/CKandCK+.htm`.

3.1.4 Técnicas utilizadas

Diferentes técnicas computacionais têm sido aplicadas para realizar a análise de EFs. As mais frequentes são técnicas capazes de gerar modelos de classificação (baseadas em aprendizado de máquina supervisionado): *Support Vector Machine* (SVM) foi a técnica mais utilizada dentre os artigos encontrados (31%), seguida por *Bayesian network* (14%)

e *Nearest Neighbour Classifier* (11%). O restante desta seção é dedicado a descrever brevemente como cada um dos artigos aplicou as estratégias de análise.

A SVM tradicional foi aplicada em Huang e Lin (2008) para fazer a comparação de cada EF sob análise com a EF neutra. O vetor de características foi composto por informações de posição (x, y), ângulo e distância entre pontos, extraídas de 19 pontos da face. Com o uso desta representação, as SVM modeladas alcançaram uma acurácia média de 81,5%.

Uma estratégia similar foi aplicada em (SARVADEVABHATLA et al., 2011). A SVM tradicional também foi utilizada, contudo, para a implementação da decisão de classificação final, a estratégia um contra todos foi aplicada para que a análise de sete classes (sete tipos de EF) pudesse ser realizada. O vetor de características utilizado contou com informações de textura de regiões da imagem da face, obtida a partir da criação de uma grade para dividir as regiões, e da aplicação de *Local Binary Patterns* (LBP). As SVMs combinadas a informações de textura alcançaram uma acurácia média de 97,6% analisando as expressões da base de dados Cohn-Kanade.

Wang et al. (2010) utilizaram uma quantidade maior de pontos (22 pontos) extraídos das imagens das faces com o apoio da *Luxand faceSDK*[2]. Esses pontos foram utilizados na construção de dois vetores de características, sendo o primeiro formado por oito ângulos extraídos dos pontos que representam a face, e o segundo vetor composto pela distância euclidiana entre os pontos da face neutra e os pontos da face que está sendo analisada. Estes dois vetores foram utilizados como entrada para uma SVM multi-classes e alcançou uma acurácia média de 87,5%.

Lemaire et al. (2011) utilizaram algoritmos de aprendizado não supervisionado para idenficar quais regiões da face seriam analisadas na criação dos modelos de análise das expressões. Estas regiões foram localizadas aplicando *Statistical Facial Feature Model* (SFAM) proposto em (MPIPERIS; MALASSIOTIS; STRINTZIS, 2008), sendo identificadas dez regiões de interesse, das quais foram extraídos 19 pontos. O vetor de características para representar uma expressão foi, então, formado pelas posições (x, y) dos pontos. As expressões foram enquadradas em modelos chamados *Actions Units* (AU) que representam o movimento de um músculo ou de um grupo de músculos de uma determinada região da face. Após identificação e representação das regiões responsáveis pela descrição das EFs, Lemaire et al. (2011) utilizaram uma SVM multi-classes para classificação de seis

[2] http://www.luxand.com/facesdk/

expressões considerando quatro níveis de intensidade, obtendo uma acurácia média de 75,8%.

Aplicam SVM na classificação de EFs também os autores de: (DAHMANE; MEUNIER, 2012), (ZHANG; TJONDRONEGORO; CHANDRAN, 2011) e (RUDOVIC; PATRAS; PANTIC, 2010). O primeiro faz uso de informação extraída de 33 regiões da face como entrada para a SVM. A informação das regiões obtida em uma face com expressão neutra é subtraída da informação das regiões referentes a uma expressão diferente; já o segundo usa as próprias imagens como entrada para a SVM (a região de interesse na imagem é encontrada por meio da aplicação do algoritmo de *Viola-Jones*; e por fim, o terceiro aplica análise discriminante linear (LDA – Linear Discriminant Analysis) para extração das características que comporão o vetor de dados de entrada para o classificador.

Vale salientar que o trabalho de Zhang, Tjondronegoro e Chandran (2011) teve o objetivo de lidar com o problema de oclusões, compondo um ambiente mais complexo de análise do que outros trabalhos. A acurácia média obtida por esses autores foi de 75%. Também é interessante destacar que no trabalho de Rudovic, Patras e Pantic (2010) foram realizados experimentos com poses frontais (onde a acurácia média obtida foi de aproximadamente 75%[3]), não frontais (com acurácia média de aproximadamente 74%) e desconhecidas (com acurácia média de 73%).

Finalmente, o último artigo analisado que faz uso de SVM aplica uma variação da técnica chamada *GentleSVM* (WU; SHEN; FU, 2011). Na realidade, a GentleSVM é o uso de SVMs dentro de uma técnica de *boosting*, que tem o objetivo de melhorar a construção de classificadores que juntos resolverão um determinado problema. As características que descrevem as EFs foram obtidas por meio da aplicação de filtros de Gabor e a acurácia obtida na tarefa de classificação das expressões foi de 85,42%.

Uma variação da SVM tradicional foi aplicada, como é o caso de (WU; SHEN; FU, 2011). *Support Vector Clustering* foi aplicada em (ZHOU; LIANG; ZHU, 2008) a fim de construir um modelo de agrupamento para EF. Nesse trabalho, somente a região de um dos olhos e da boca são usadas para extração das características, que são obtidas por meio de filtros de Gabor e compoem um vetor de 94 características.

[3] Essas medidas foram calculadas pelo autor deste trabalho com base nas acurácias médias obtidas para cada uma das EFs que representam as seis emoções universais e a expressão neutra.

A estratégia de *boosting* foi também usada por Zhao et al. (2011), porém com classificadores implementados via *Hidden Conditional Random Fields* (HCRFs). Nesta estratégia os autores conseguem uma acurácia de 96,6% na classificação das EF.

Modelos Escondidos de Markov (HMM – *Hidden Markov Models*) foram aplicados como modelos de classificação em dois trabalhos: (POPA; ROTHKRANTZ; WIGGERS, 2010) e (HOEY; LITTLE, 2007). Popa, Rothkrantz e Wiggers (2010) optaram por usar 13 regiões de interesse dentro da face, usando AAM para realização da subtração das coordenadas de pontos (x, y) entre dois *frames* consecutivos. O uso desse modo de representação das características da face combinado ao uso de HMM resultou em uma acurácia média de 93% para 21 modelos de expressões. Já os autores de (HOEY; LITTLE, 2007) utilizam *optical flow* para extração das características que são usadas junto de modelos *coupled hidden Markov models* (CHMMs) em uma análise temporal de *frames* de vídeo. Nesse trabalho *optical flow* é usado para encontrar as regiões da face, as quais são representadas em dois vetores de características com 16 e 32 posições. A acurácia média obtida foi de 80%.

Valenti, Jaimes e Sebe (2008), Joho et al. (2009) e Valenti, Jaimes e Sebe (2010) utilizaram a estratégia *Piecewise Bezier Volume Deformation* (PBVD) (TAO; HUANG, 1998) para seguir 16 pontos das EFs fornecidos por um modelo em 3D. Um vetor de 12 características foi montado com a direção e intensidade com que esses pontos se movimentavam para posteriormente serem classificados em uma Rede Bayesiana.

Os autores de (YONG; SUDIRMAN; CHEW, 2011) utilizaram a aplicação *FaceGen Modeller* (BALOMENOS et al., 2005) para gerar as EFs em um avatar, e destes avatares foram extraídos quatro regiões da face (sobrancelhas, olhos, nariz e boca) que foram submetidos em um algoritmo linear de extração das componentes principais de cada região (PCA). A partir destas características, os autores usaram um classificador Baysiano que obteve uma acurácia média de 90,6%.

Redes Neurais Artificias também aparecem como estratégias de implementação de classificadores nos trabalhos analisados. Os autores de (PETRIDIS; PANTIC, 2011) utilizaram 20 pontos (x, y) da face que foram inicialmente marcados manualmente e seguidos por meio do uso do filtro *Patras-Pantic* (PATRAS; PANTIC, 2004). Esses pontos foram analisados por meio de PCA a fim de selecionar as informações que melhor descreviam o movimento da cabeça e, em seguida, os pontos localizados na boca foram usados como referência para acompanhar o movimento total da face. Vale ressaltar que este projeto possuia como objetivo diferenciar a fala de uma risada e utilizou informações multimodais (provenientes

de sons e imagens) como entradas de uma rede neural *feedforward*. A estratégia obteve uma acurácia de 83,9% no reconhecimento do movimento da boca, 53,2% para o movimento da cabeça e 82,3% para ambos. Uma rede neural treinada com algoritmo de retro-propagação do erro foi aplicada por Yu, You e Tsai (2012), a partir de uma representação da EF que utilizava 12 características extraídas de 19 pontos da face. No entanto, houve um baixo desempenho (média de 50% de acertos). Já em (CHANG; HUANG, 2010), os autores normalizam as faces em 200x250 pixels, utilizam 17 distâncias entre 16 pontos da face, além de 16 regiões da face. PCA foi utilizado para extração das características principais de cada face, sendo que foi construído um modelo para cada pessoa, tendo como base a face neutra. Então, a *Radial-Basis Function Neural Network* (RBFNN) foi usada para realizar a classificação das expressões.

Guo e Ruan (2011) utilizam matrizes binárias de covariância para detecção das características da face, utilizando um *threshold* para transformar a imagem em uma matriz binária, obteve-se uma acurácia máxima de 95,24%, utilizando *local binary covariance matrices* (LBCM), apesar de citar que não é muito viável sua utilização devido ao alto custo computacional.

Lajevardi e Hussain (2010) utilizam Trasformada Rápida de Fourier (FFT) nos *frames* que são posteriormente tratados com Filtros de *Log-Gabor* e são submetidos à transformada inversa de Fourier, sendo obtidos nesse processo 24 amplitudes que serão utilizadas na classificação da expressão por LDA. Este projeto teve uma acurácia de 82,5% para imagens com resolução 32x32 *pixels* e uma melhora de menos que 2% para imagens com resolução 64x64 *pixels*.

Li, Ruan e Li (2010) utilizam *single-tree complex wavelet transform* (ST-CWT) para extração das características que foram submetidas a um algorítmo de PCA para redução da redundância nos dados de entrada, classificadas por distância euclidiana, com uma acurácia de 88,6% para a base de dados JAFFE e 96,83% para base de dados CK.

Liu, Ruan e Wang (2011) trabalham com reduções não lineares das características para classificação por meio de distância euclidiana, o autor testa sua técnica com os algoritmos OTR1DGE, PCA, LDA, LPP, NPE, OLPP e TSA, aplicando dois treinamentos (40 e 80 características) por expressão e obteve uma melhora significativa no segundo teste, destacando-se PCA que passou de 80,74% para 94,56% em sua taxa de reconhecimento.

Zhi e Ruan (2008) utilizam *supervised spectral analysis* (SSA), motivado por clusterização expectral, para extrair as características das EFs e classificá-las.

Em (ZHANG; GENG, 2011), os autores extraíram regiões da face (olhos, região entre os olhos, boca e nariz) de forma manual e, na sequência, aplicaram PCA e LDA para extrair características descritivas dessas regiões. Essas características foram então submetidas a classificadores KNN (*Nearest Neighboor Classifiers*) para classificação das expressões. Vale ressaltar que o autor destaca a influência de cada região no reconhecimento das expressões, por exemplo, o uso da região da boca levou a uma taxa de reconhecimento de 81,82% para expressões de raiva e a uma taxa de reconhecimento de 30,91% para expressões de desgosto. Os autores ainda usaram um vetor de características composto pela subtração das imagens com expressões de emoções, das imagens com expressões neutras e, para essa estratégia, a taxa de reconhecimento considerando o mesmo contexto subiu respectivamente para 83,64% e 55.45%.

Classificadores KNN são também aplicados por Wang et al. (2011) e Buenaposada, Muñoz e Baumela (2008). Em (WANG et al., 2011) é feito um estudo com algumas variações de LDA, com a finalidade de verificar o efeito de diferentes extratores de características. Assim, diferentes estratégias são usadas para extrair características e todas as representações obtidas são usadas no KNN. As extratégias aplicadas são: *linear discriminant analysis* (LDA), *local fisher discriminant analysis* (LFDA), *linear boundary discriminant analysis* (LBDA), *linear boundary discriminant analysis* (RNBDA). As melhores taxas de reconhecimento das expressões são obtidas usando RNBDA com uma taxa de 81,43%. Em (BUENAPOSADA; MUÑOZ; BAUMELA, 2008), os autores usam LDA e *Locality Preserving Projections* (LPP) para extrair as características das imagens e então usar o classificador KNN. O trabalho focou em estudar imagens com diferentes luminosidades e alcancou uma acurácia média de 89%.

Cho e Park (2011) trabalham com PCA e ICA para detecção de características nos próprios pixels das imagens que foram subtraídas de uma referência incial. O trabalho foca no reconhecimento de faces e direção da cabeça, utilizando distância euclidiana em sua classificação.

Xiang, Leung e Cho (2008) primeiramente utiliza a distância entre os olhos como referencia inicial para encontrar a região de interesse em uma face e, em seguida, aplica Fourier para extração das informações temporais na movimentação dos *pixels*. Este foi o único artigo analisado que utilizou Teoria de Conjuntos Fuzzy no modelo de análise, e obteve 88,8% de acurácia média no reconhecimento das EFs.

Tews et al. (2011) utilizam dez pontos na face que eles consideram como os pontos mais relevantes em relação aos músculos faciais. Os autores também descrevem as áreas (entre os pontos) que foram utilizadas nos vetores de características e ainda relatam que a região da boca é a que mais colaborou com a classificação das expressões, utilizando somente as medidas das áreas de expressões rotuladas como parâmetro para classificação.

Yang e Chiang (2008) trabalham com síntese de movimentos e utiliza distâncias e ângulos para extração das características de imagens reais.

3.1.5 Metodologias de Avaliação de Desempenho

A Tabela 3 mostra quais estratégias foram utilizadas para estimar a acurácia dos modelos apresentados nos artigos analisados nesse levantamento bibliográfico. Nove trabalhos não utilizaram e/ou não citaram nenhuma estratégia e portanto não compõem a referida tabela.

Tabela 3 – Técnicas para Validação dos Modelos

Estratégia	Estudo
Holdout	(LI; RUAN; LI, 2010) (LAJEVARDI; HUSSAIN, 2010) (HUANG; LIN, 2008) (SIDDIQUI; LIAO; MEDIONI, 2009) (XIANG; LEUNG; CHO, 2008) (LAJEVARDI; HUSSAIN, 2010) (YONG; SUDIRMAN; CHEW, 2011) (ZHOU; LIANG; ZHU, 2008) (SONG et al., 2008)
Leave one out	(HOEY; LITTLE, 2007) (ZHANG; GENG, 2011) (GUO; RUAN, 2011) (ZHI; RUAN, 2008) (WU; SHEN; FU, 2011) (CHANG; HUANG, 2010)
k-fold Cross-validation	(PETRIDIS; PANTIC, 2011) (ZHANG; TJONDRONEGORO; CHANDRAN, 2011) (DAHMANE; MEUNIER, 2012) (SARVADEVABHATLA et al., 2011) (WANG et al., 2010) (LEMAIRE et al., 2011) (POPA; ROTHKRANTZ; WIGGERS, 2010) (BUENAPOSADA; MUÑOZ; BAUMELA, 2008) (RUDOVIC; PATRAS; PANTIC, 2010)
Bootstrap	(LIU; RUAN; WANG, 2011) (CHEN et al., 2012)

Para avaliar e comparar o desempenho das técnicas utilizadas nos projetos, os artigos (SARVADEVABHATLA et al., 2011), (LEMAIRE et al., 2011), (ZHAO et al., 2011), (PETRIDIS; PANTIC, 2011), (LAJEVARDI; HUSSAIN, 2010), (YONG; SUDIRMAN; CHEW, 2011), (CHANG; HUANG, 2010) e (BUENAPOSADA; MUÑOZ; BAUMELA, 2008) utilizaram "Matrizes de Confusão". Através das matrizes apresentadas, pode-se perceber que a EF de medo é a que mais se confude com as outras pois os pares de expressões (medo e triste), (medo e

feliz), (medo e nojo) e (raiva e nojo) são os pares que apresentaram maior índice de erros nas classificações dos respectivos artigos. O artigo (ZHAO et al., 2011), que analisa a EF das pessoas enquanto assistem vídeos para classificação automática do gênero dos vídeos, discutiu a dificuldade inerente à classificação das EFs relacionadas a vídeos de tragédia e tristeza.

Ainda com relação ao desempenho das técnicas, os artigos (DAHMANE; MEUNIER, 2012) e (CHANG; HUANG, 2010) mostram uma particularidade que deve ser considerada ao pensar em um produto comercial, pois ao analisar os algoritmos deixando uma pessoa fora da etapa de treinamento e validação, a acurácia das técnicas caem significativamente, de 95,71% para 63,51% em (DAHMANE; MEUNIER, 2012) e 95,9% para 78,1% em um dos testes realizados em (CHANG; HUANG, 2010). Ou seja, estes fatos demonstraram que essas ferramentas teriam baixo desempenho em aplicações com indivíduos novos, cujos dados não fizeram parte do treinamento dos modelos.

3.2 Expressões Faciais Gramaticais

O levantamento bibliográfico organizado nesta seção diz respeito a estudos conduzidos na área de reconhecimento das EFs considerando o escopo das LSs. A partir da pesquisa aqui descrita, percebe-se que as análises realizadas nessa área possuem como objetivo principal auxiliar o reconhecimento dos sinais das LSs em estudos multimodais. Usar informações sobre as EFs ajuda a retirar as ambiguidades e aumenta a acurácia das técnicas envolvidas nos projetos de reconhecimento.

3.2.1 Escopo dos Estudos

Não foi encontrado nenhum artigo que abrangesse todas as EFs usadas em um contexto de LS. Além disso, é notável que muitos trabalhos executam análises cujo objetivo é auxiliar o reconhecimento de apenas um único sinal (KELLY et al., 2009a).

Uma exceção a esse escopo mais restrito é o trabalho de Kacorri (2013). Nessa pesquisa, o autor modela as sentenças em *American Sign Language* (ASL) e mostra como essas expressões podem ocorrer durante a sinalização, além de poderem ocorrer simultaneamente, como pode ser visto na Figura 12. As EFGs podem não modificar o significado semântico do sinal executado, mas influenciam de maneira semântica as orações:

uma frase afirmativa, por exemplo, passa a ser interrogativa sem modificar nenhum sinal, porém modificando a EFG.

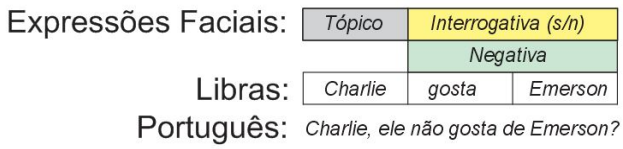


Figura 12 – Modelagem de uma sentença em LS usando EFGs

Os sinais originais realizados no exemplo de Kacorri (2013) são “CHARLIE”, “LIKE” e “EMERSON”. Essa oração sem o uso de EFGs, teria um sentido afirmativo (Charlie gosta de Emerson). Com a utilização somente da EF Interrogativa (s/n), ela passa a significar “Charlie gosta de Emerson?”, e finalmente, da maneira que foi mostrada na Figura 12, com a EFG de tópico durante a sinalização de Charlie e a EFG interrogativa junto a EFG de negação no restante da oração, ela passa a significar “Charlie, ele não gosta de Emerson?”.

3.2.2 Natureza dos Dados e Pré Processamento

De maneira resumida, pode-se dizer que os autores dos artigos analisados estão preocupados com a extração de características de todas as regiões da face de imagens com ruídos (DING; MARTINEZ, 2010) ou de somente parte das regiões, como (SAEED, 2010) que analisa EFGs usando apenas a extração das características dos lábios. Outros discutem uma modelagem que precisa considerar o domínio do tempo e a simultaneidade das expressões (KACORRI, 2013), (KOSTAKIS; PAPAPETROU; HOLLMÉN, 2011) e (CARIDAKIS; ASTERIADIS; KARPOUZIS, 2011). Alguns dos estudos analisados tratavam da análise do reconhecimento manual dos sinais levando em consideração os sinais não-manuais em uma análise multimodal (KELLY et al., 2009b), (KRNOUL; HRUZ; CAMPR, 2010), (AGRIS; KNORR; KRAISS, 2008), (ARI; UYAR; AKARUN, 2008), (YANG; LEE, 2011), (NGUYEN; RANGANATH, 2012),(MICHAEL; METAXAS; NEIDLE, 2009) e (KELLY et al., 2009a).

Todos os estudos analisados no escopo deste levantamento utilizaram vídeos em suas análises, e isso se deve ao escopo temporal do problema, já que as EFGs e outros elementos da língua fazem sentido na execução de uma sequência de ações, que podem ocorrer de maneira sequencial e/ou de maneira simultânea. A Tabela 4 mostra uma classificação dos

artigos com relação à abordagem utilizada para representação da informação (textura ou pontos geométricos), tipos de dados (primários e secundários) e condições dos dados (obtidos em laboratório ou em um ambiente externo), características essas que já foram explicadas na Seção 3.1.2.

Tabela 4 – Informações sobre os dados utilizados nas experimentações e aplicações dos estudos referentes às EFGs.

Estudo	Abord.		Tipos		Condições	
	Pix.	Geo.	Pri.	Sec.	Lab.	Ext.
(ARI; UYAR; AKARUN, 2008)	x	x	-	x	x	-
(NGUYEN; RANGANATH, 2008)	x	x	x	-	-	-
(AGRIS; KNORR; KRAISS, 2008)	-	-	-	x	x	-
(KELLY et al., 2009a)	x	-	x	-	x	-
(MICHAEL; METAXAS; NEIDLE, 2009)	x	-	-	x	x	-
(DING; MARTINEZ, 2010)	x	x	-	x	x	-
(KELLY et al., 2009b)	x	-	x	-	x	-
(KRNOUL; HRUZ; CAMPR, 2010)	x	-	-	x	x	-
(SAEED, 2010)	x	x	x	-	x	-
(CARIDAKIS; ASTERIADIS; KARPOUZIS, 2011)	x	-	-	-	-	x
(KOSTAKIS; PAPAPETROU; HOLLMÉN, 2011)	x	x	-	x	-	-
(YANG; LEE, 2011)	x	x	-	x	x	-
(NGUYEN; RANGANATH, 2012)	x	-	-	x	-	-
(KACORRI, 2013)	-	-	-	-	x	-

Outra preocupação presente nas pesquisas nessa área é com relação à oclusão da face durante a captura dos sinais, pois muitos sinais devem ser executados em frente a face, prejudicando a aquisição dos dados que descrevem uma EF. Nguyen e Ranganath (2008) e Ding e Martinez (2010) utilizam interpolação para os casos de oclusão, usando como referência os dados capturados antes do momemento em que a oclusão da face ocorre.

3.2.3 Bases de Dados

Os autores Nguyen e Ranganath (2008), Kelly et al. (2009a), Kelly et al. (2009b) e Saeed (2010) não citaram bases de dados em seus trabalhos e possivelmente utilizaram dados primários. Já Ding e Martinez (2010) utilizou a base de dados AR (MARTINEZ, 1998), a qual possui quatros EFs diferentes (neutro, feliz, bravo e grito) e XM2VT (MESSER et al., 1999) que possui 295 mil vídeos especialmente gravados para suportar o estudo de problemas de reconhecimento de pessoas em sistemas de segurança.

Os autores (KRNOUL; HRUZ; CAMPR, 2010) utilizam a base de dados UWB-07-SLR-P (CAMPR; HRÚZ; TROJANOVÁ, 2008) que possui dados gravados em ambiente de laboratório com fundo preto e iluminação homogênea, com 378 sinais gravados, cinco ou mais vezes com três perspectivas diferentes.

Já Kostakis, Papapetrou e Hollmén (2011) utilizaram a base de dados da *National Center for Sign Language and Gesture Resources*[4] com 873 enunciados em ASL. E Agris, Knorr e Kraiss (2008) e (MICHAEL; METAXAS; NEIDLE, 2009) utilizaram a base de dados(AGRIS; KRAISS, 2007) com vídeos captados em condições de laboratório contendo 135 sinais isolados e 780 sentenças completas executadas na LS alemã.

Ari, Uyar e Akarun (2008) utilizaram a base de dados de Aran et al. (2007) que possui 8 classes de SMN de 11 pessoas diferentes, com aproximadamente 2 segundos de vídeo para cada gravação. E finalmente, Michael, Metaxas e Neidle (2009) utilizaram a base de dados *Boston University American Sign Language Linguistic Research Project* com 42 vídeos.

3.2.4 Técnicas utilizadas

A aplicação de PCA se destacou no conjunto de artigos analisados. Em 9 dos 14 artigos, essa técnica é usada como extrator das características principais da face, evitando redundância e consequentemente diminuindo o custo computacional das técnicas utilizadas para análise das expressões (AGRIS; KNORR; KRAISS, 2008). Muitos trabalhos utilizaram modelos baseados em PCA, como *Active Shape Model* (ASM) (5 artigos) e *Active Appearance Model* (AAM) (6 artigos). Esses modelos são utilizados para encontrar um contorno correspondente à face, sendo AAM uma forma generalizada da ASM, acrescentando a informação de textura da imagem.

A velocidade e direção dos movimentos da cabeça são as características usadas por Aran et al. (2009) para executar análise multimodal no reconhecimento dos sinais. A análise da informação sobre os gestos não-manuais (as EFGs) foi usada como informação complementar à análise de sinais manuais, e foram realizadas por meio da aplicação de HMM, obtendo uma acurácia média de 97,8%. A técnica HMM em análise multimodal também foi aplicada por Kelly et al. (2009a). Nessa análise, os autores consideraram somente o movimento da cabeça como sinal não-manual e utilizaram a medida entre os

[4] `http://www.bu.edu/asllrp/cslgr/`

olhos como referência para calcular o movimento da cabeça *frame* a *frame*. Já em (KELLY et al., 2009b), os autores acrescentam as características das sobrancelhas em sua análise realizada com HMM.

Agris, Knorr e Kraiss (2008) trabalharam com 50 pontos da face para extração de sete características utilizadas como entrada em uma HMM. Nesta abordagem, os autores obtiveram uma acurácia de 80,2% e 96,9% em sinais isolados e reconhecimento para pessoas diferentes e a mesma pessoa respectivamente. Essa acurácia cai para 65,1% e 87,5% para análise em um discurso.

As técnicas HMM e SVM são usadas de forma combinada no trabalho de Nguyen e Ranganath (2012). No trabalho desses autores, 21 pontos da face, marcados manualmente, são usados para representar a informação a ser usada para classificar as EFGs. Tal abordagem alcancou uma acurácia média de 80,9%. E Michael, Metaxas e Neidle (2009) combinou PCA e SVM em uma abordagem para respectivamente reduzir informação redundante e classificar as EFGs (interrogativas e negativas), obtendo 95% de acurácia média.

Autores dos trabalhos (YANG; LEE, 2011) e (ARI; UYAR; AKARUN, 2008) aplicaram apenas SVM para classificação das EFGs. Yang e Lee (2011) utilizam 31 pontos da face para gerar 16 medidas para compor o vetor de características usado como entrada na SVM. A magnitude e a direção do movimento foi usada por Ari, Uyar e Akarun (2008) para compor as oito características em seu vetor de representação dos dados, analisados por uma SVM, com uma acurácia de 67,1% para identificação das expressões afetivas.

Saeed (2010) compara técnicas utilizadas para extração das características da boca para auxiliar no reconhecimento das EF. O vetor de características é composto por: área; o comprimento do eixo maior e menor; a excentricidade; orientação; e o comprimento do perímetro do contorno exterior da boca. Os autores utilizaram *optical flow* com 300 características dos *frames* para compor seu vetor de características.

Krnoul, Hruz e Campr (2010) utilizaram 19 pontos da face para extrair informações como: rotação da face; movimentos verticais e horizontais da cabeça; olhar; boca aberta em duas catergorias diferentes; boca comprimida; e piscar dos olhos.

Ding e Martinez (2010) utilizaram AdaBoost como estratégia em sua classificação das EF, sendo que o modelo classificador recebe as formas e texturas da imagem como entrada. Este trabalho possui como objetivo utilizar “contextos” nas imagens, i.e., imagens

borradas ou imagens com somente parte do objeto de classificação para aumentar a acurácia de seu classificador.

Caridakis, Asteriadis e Karpouzis (2011) utilizaram uma rede neural recorrente com 25 neurônios na camada de entrada e 20 neurônios na camada oculta, mas não informam como foi construído o vetor de características.

3.2.5 Metodologias de Avaliação de Desempenho

A Tabela 5 mostra quais estratégias foram utilizadas para estimar a acurácia dos modelos apresentados nos artigos analisados nesse levantamento bibliográfico. Os demais trabalhos não utilizaram e/ou não citaram nenhuma estratégia e portanto não compõem a referida tabela, lembrando que alguns trabalhos não estavam focados em reconhecimento automático, mas foram considerados por causa das informações dentro do escopo deste trabalho.

Tabela 5 – Técnicas para Validação dos Modelos

Estratégia	Artigos
Holdout	(NGUYEN; RANGANATH, 2008) (SAEED, 2010) (AGRIS; KNORR; KRAISS, 2008) (YANG; LEE, 2011) (HRÚZ; TROJANOVÁ; ŽELEZNÝ, 2011)
k-fold Cross-validation	(ARAN et al., 2009) (ARI; UYAR; AKARUN, 2008)

Para avaliar e comparar o desempenho das técnicas utilizadas nos projetos, os artigos (KELLY et al., 2009a), (KELLY et al., 2009b), utilizaram Clusterização iterativa que pode ser visto em detalhes em (KELLY; MCDONALD; MARKHAM, 2009). Ari, Uyar e Akarun (2008) utilizou matrizes de confusão e identificou maiores problemas entre as expressões afirmativas e "afirmativas felizes"e "felizes e tristes"com uma acurácia geral em torno de 67,1%, no entanto destaca-se com resultados em torno de 95% as expressões para perguntas.

O trabalho (NGUYEN; RANGANATH, 2012) que possui uma análise mais próxima das expressões faciais estudadas neste trabalho, obteve uma acurácia 91,76% utilizando um algoritmo em que a marcação inicial precisa ser feita de maneira manual para iniciar um processo de predição baseada em formas geométricas e em somente uma pessoa, para

testes realizados com mais de uma pessoa, a técnica obteve uma acurácia em torno de 87,7% para expressões faciais gramaticais da língua de sinais americana.

Os demais trabalhos realizaram reconhecimento automático de expressões faciais com intuito de auxiliar na tradução de sinais manuais, ou não apresentaram experimentos que possuissem escopo parecido com esta dissertação de mestrado.

3.3 Considerações Finais

Este capítulo apresentou uma revisão na literatura com escopo nas EFAs e EFGs, com objetivo de identificar as principais técnicas de normalização, extração de características, representação e importância da informação temporal e principais técnicas de reconhecimento utilizadas. Tais estudos foram base para a construção deste trabalho, bem como identificação dos principais desafios que seriam encontrados ao longo desta dissertaçao de mestrado.

Através do estudo realizado na área de reconhecimento de EFAs e EFGs, constata-se que no contexto de análise de EFGs existe uma clara preocupação, maior do que no caso de análise de EFAs, com problemas relacionados a dependências temporais e com oclusões. No primeiro caso, a importância da representação temporal está ligada ao fato da execução da expressão facial estar presente na sinalização de um ou mais sinais. Já o problema de oclusões, caracterizado por perda da informação proveniente, muitas vezes, da presença da mão entre o sensor de captura e a face, representa um fator de análise imprescindível para o caso de aplicações onde o objetivo final é o reconhecimento automático de sinais.

4 Aprendizado de Máquina com Perceptron Multicamadas

A capacidade de um programa de computador de utilizar dados históricos para melhorar seu desempenho na resolução de determinada tarefa, é definida por Mitchell (1997) como Aprendizado de Máquina. Este contexto está relacionado ao aprendizado indutivo, no qual o sistema tenta generalizar uma solução por meio de dados *a priori*. Nesse processo é utilizada uma medida de erro para auxiliar a dinâmica do "aprendizado", já que em cada ciclo do processo, o desempenho do programa é medido para verificar a necessidade de alterar seus parâmetros com objetivo de encontrar uma solução melhor para a tarefa em questão.

As Redes Neurais Artificiais (RNAs), geralmente chamadas de "redes neurais", buscam simular a forma como o cérebro humano processa as informações, e sua capacidade, muitas vezes, está associada a um grande número de neurônios interconectados por meio de sinapses. As sinapses são responsáveis por possibilitar a realização da tarefa para a qual a rede neural está sendo construída, recebendo, processamento e armazenando informação. Exemplos de problemas típicos para estas estruturas computacionais são o reconhecimento de padrões e aproximação de funções.

Assim como o cérebro humano, as RNAs possuem algumas características úteis para a solução de problemas: processamento neuronal linear ou não-linear, sendo que a não-linearidade presente nessas estruturas possibilitam que ela generalize informações complexas; mapeamento entre as informações de entrada e saída, realizado por meio da modelagem de sinapses (pesos de conexões) que se adaptam ao conhecimento a ser adquirido, importante para problemas que possuem uma dinâmica com relação ao tempo, como por exemplo, problemas envolvendo a análise dos dados da bolsa de valores; por fim, vale ressaltar sua analogia ao sistema biológico, o cérebro humano, responsável por resolver problemas lineares e não-lineares e servir de inspiração para o estudos das RNAs com o objetivo de colaborarem na solução de problemas reais, tais como o apresentado neste trabalho de mestrado.

O conceito de Aprendizado de Máquina pode ser associado ao contexto das RNAs, pois utiliza-se as redes neurais como forma de modelar o conhecimento que é adquirido pelo programa durante as iterações de aprendizado. Existem outras abordagens para o aprendizado indutivo, como Árvores de Decisão e Máquinas de Vetores de Suporte (*Support Vector Machines* - SVM). Este trabalho foi desenvolvido dentro do contexto das

redes neurais, com foco nas redes Perceptron Multicamadas (MLP). Além de simular o cérebro humano, as MLPs são ótimos detectores de características, devido suas camadas ocultas que são formadas por neurônios do tipo Percetron interconectados, responsáveis por realizarem localmente, e de forma eficiente, a discretização do erro envolvido na tarefa de aprendizado.

Assim, esse capítulo é dedicado a apresentar os conceitos fundamentais sobre MLP e está organizado da seguinte forma: a primeira Seção 4.1 apresenta um panorama geral da técnica MLP, a segunda Seção 4.2 apresenta uma metodologia de treinamento que é amplamente utilizada e a terceira Seção 4.3 descreve os parâmentros que foram trabalhados nesta dissertação de mestrado.

4.1 Perceptron Multicamadas

As redes neurais Multilayer Percetrons (MLPs) surgiram a partir do conceito básico introduzido por Rosenblatt em 1958, chamado de Percetron, capaz de resolver problemas simples com padrões linearmente separáveis. Em MLPs há uma rede de Perceptrons com alto grau de conectividade, que pode ter uma ou mais camadas de neurônios internos entre a entrada e a saída da rede, sendo que cada neurônio possui uma função de ativação não-linear, diferenciável em todo seu domínio. Este tipo de rede neural teve seu potencial destacado após os estudos de Rumelhart e McClelland em 1986, com o algoritmo de retropropagação do erro, responsável pelo treinamento e ajuste dos pesos de cada neurônio de acordo com a retropropagação da informação de erro encontrada em cada iteração do treinamento.

Um Perceptron pode ser formalmente definido pela equação 1, onde w é o vetor de peso aplicado nas entradas do neurônio, x é o vetor de entrada, b é um fator de bias que pode ser positivo ou negativo e que é ajustado a cada ciclo de aprendizado juntamento com os vetores de pesos, e φ é a função de ativação que será responsável pela resposta do neurônio dada uma entrada. Assim, a equação que define um Perceptron é

$$y = \varphi\left(\sum_{i=1}^{n} w_i x_i + b\right) = \varphi\left(\mathbf{w}^T\mathbf{x} + \mathbf{b}\right), \tag{1}$$

onde n é a dimensão da entrada fornecida ao neurônio.

As redes MLPs são compostas por neurônios (Perceptrons) interligados entre si, e geralmente são organizadas da seguinte maneira: uma camada de entrada responsável por receber a informação que será processada pelos neurônios da rede; uma ou mais camadas escondidas, que são os grupos de neurônios entre a camada de entrada e a camada de saída; e finalmente, a camada de saída que é responsável por produzir a resposta da rede. Uma arquitetura genérica de uma rede MLP pode ser vista na Figura 13, na qual os circulos representam os neurônios, os quadrados representam as unidades de entrada e as linhas conectando os neurônios representam as sinapses que ponderam os sinais enviados aos neurônios.

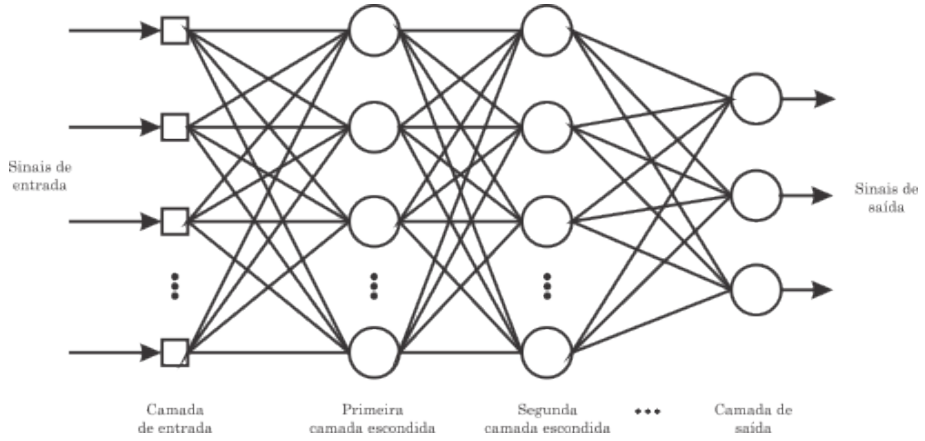


Figura 13 – Exemplo de arquitetura de uma rede neural MLP (HAYKIN et al., 2009)

A propagação da informação acontece de forma progressiva neste tipo de rede, i.e., cada neurônio é conectado a todos os neurônios da camada seguinte, com a informação passando de camada em camada, sem a presença de realimentação. Para este tipo de rede neural, se faz necessário a utilização de um algoritmo de aprendizagem, responsável por ajustar os pesos das conexões dos neurônios, e consequentemente por extrair as características do problema. Geralmente este algoritmo atua de duas formas: de forma sequencial (*online*), na qual o erro é analisado a cada iteração, ou em lote (*batch*), com o erro sendo calculado somente ao final de uma época, i.e., somente depois que todos os dados de treinamento forem apresentados à rede.

O algoritmo baseado em lote é caracterizado por possuir uma forma simples de encontrar um mínimo local, utilizado geralmente em problemas de regressão não-linear. Neste trabalho, utilizou-se o Algoritmo de Retropropagação do Erro (*Backpropagation*), em sua forma sequencial. Este algoritmo é caracterizado por ser simples de implementar e por apresentar eficiência na solução de problemas de alta complexidade em classificações

de padrões, proporcionado pela capacidade de generalização que ocorre pela correção do erro a cada iteração.

4.2 Algoritmo de Retroprogação do Erro

O algoritmo de Retropropagação do Erro (*Backpropagation*) é um algoritmo comum em arquiteturas de redes neurais com aprendizado supervisionado. Primeiramente foi descrito por Paul Werbos em sua tese de doutorado (WERBOS, 1974) e redescoberto por McClelland et al. (1986), possibilitando um grande avanço na área de redes neurais (HAYKIN, 2009).

De maneira resumida (para mais informações, consulte (HAYKIN, 2009)), deve-se determinar o erro numa iteração m, que consiste na subtração entre o valor desejado $d_j(m)$ pelo valor resultante da iteração $y_j(m)$. A soma do erro instantâneo quadrático de uma interação s, representado por ϵ_{av}, será responsável por representar a função custo 2, que de maneira inversa, é o desempenho da rede mensurado por uma função de ativação. Essa função custo representa a necessidade de ajustes nos parâmetros da rede neural para a obtenção do aprendizado (aproximação) de um padrão desejado.

$$\epsilon_{av} = \frac{1}{M} \sum_{m=1}^{M} \epsilon(m) \tag{2}$$

O algoritmo de Retropropagação do Erro aplica uma correção Δw_{ji} nos pesos sinápticos, sendo i o número de valores de entrada em um neurônio e j o identificador do neurônio. A correção é proporcional à derivada parcial do somatório dos erros encontrados na última camada $\epsilon(m)$ com relação ao Δw_{ji}, i.e., $\partial\epsilon(m)/\Delta w_{ji}$, que definirá qual a taxa que deverá ser aplicada ao peso Δw_{ji} à medida que o erro $\epsilon(m)$ se modifica.

O algoritmo básico de treinamento de uma rede neural MLP é resumidamente descrito no algoritmo **??**. A formulação completa do processo de minimização do erro usada neste algoritmo pode ser encontrada em (HAYKIN, 2009).

4.3 Estudo dos Parâmetros do Perceptron Multicamadas

De acordo com o problema que pretende-se resolver, alguns parâmetros da rede neural e do algoritmo de aprendizado aplicado a ela devem ser determinados com o intuito de obter melhores resultados. Os principais parâmetros são:

Algoritmo 1 Algoritmo de treinamento de uma MLP utilizando Retroprogacação do erro com gradiente descendente

Entrada: $x \leftarrow entrada$ ▷ A entrada do neurônio i na camada j é denotado x_{ji} e o peso do neurônio i em j é denotado w_{ji}
Entrada: $d \leftarrow rotulos$
1: Inicialização
2: Atribuição de valores aleatórios para o conjunto de vetores de pesos sinápticos w
3: Atribuição do valor inicial para taxa de aprendizado η
4: **Enquanto:** Critério de parada **Faça** ▷ Possíveis condições de parada: número máximo de épocas, erro mínimo ou teste de desempenho no conjunto de validação
5: $t \leftarrow 1$
6: **Para cada:** (x, d) **Faça**
7: Apresente o vetor de entrada x, propague-a pelas camadas da rede computando as saídas de cada neurônio j das camadas escondidas e de saída
8: Para cada neurônio da camada de saída, calcule a informação de erro (o gradiente)
9: Para cada neurônio da camada escondida, calcule a informação de erro (o gradiente)
10: Com as informações de erro, ajuste os conjunto de vetores de pesos $w_{ji}(t) = w_{ji}(t-1) + \eta * \Delta w_{ji}(t)$

- Função de Ativação: o gradiente local de cada neurônio MLP requer a derivada da função de ativação, sendo então necessário que essa função seja contínua em todo seu domínio. Duas funções são comumente utilizadas em redes MLP, sendo elas: função sigmoidal e a função tangente hiperbólica. A derivada da função sigmoidal representa uma curva na qual há uma maior alteração nos neurônios cujo os sinais assumem valores intermediários, proporcionando maior estabilidade ao sistema, ao contrário da função tangente hiperbólica, que sua derivada resulta em uma função com uma transição mais suave.
- Taxa de Aprendizado: este parâmetro é responsável por auxiliar na mudança dos pesos sinápticos. Ele pode assumir um valor constante em $[0, 1]$ ou pode ter o seu valor alterando, dentro deste intervalo de acordo com heurísticas de adaptação da taxa de aprendizado. Essa taxa é aplicada ao gradiente local no momento de atualização dos vetores de peso. Sendo assim, quanto maior for o valor desta taxa, maior a velocidade do aprendizado, no entanto, isto pode levar à uma oscilação do modelo ao redor do erro.
- Critério de parada: este parâmetro geralmente está associado ao erro médio quadrático, responsável por dizer se o erro obtido na saída da rede neural já é suficientemente pequeno para que o processo de treinamento possa ser finalizado. Vale ressaltar que

este parâmetro é subjetivo e dependendo do valor atribuído a ele e da complexidade do problema, o algoritmo pode estacionar em um mínimo local.

- Número de neurônios na(s) camada(s) escondida(s): não há regras formais determinadas para tal especificação, no entanto, sabe-se que o número de camadas escondidas, bem como o número de neurônios em cada camada escondida é responsável por extrair as características do padrão que pretende-se aprender.

4.4 Considerações Finais

Este capítulo apresentou uma breve descrição da arquitetura das RNA MLPs, do algoritmo de Retropropagação do Erro e dos principais parâmetros que devem ser ajustados em um processo de treinamento de uma MLP. Como os valores de tais parâmetros são encontrados de forma empírica, algumas combinações de valores para eles foram testadas neste trabalho com o objetivo de encontrar a melhor combinação para a resolução do problema em questão. O Capítulo 5 descreverá melhor como esses parâmetros foram explorados e o Capítulo 6 mostrará os resultados provenientes destas escolhas.

5 Reconhecimento de Expressões Faciais Gramaticais: contexto e experimentos

Como apresentado no início desta dissertação, este estudo tem como um de seus objetivos, o desenvolvimento de um conjunto de modelos de reconhecimento de padrões capazes de resolver o problema de reconhecimento de expressões faciais usadas no contexto da Libras, as Expressões Faciais Gramaticais, considerando-as em nível sintático.

Nesta dissertação de mestrado, uma expressão facial $EF_i \in \{EF_1, EF_2, ...EF_n\}$ é a forma como os pontos $\{p_1, p_2, ...p_n\}$ extraídos da face humana estão dispostos no espaço tridimensional. Estes pontos possuem coordenadas (x, y, z), sendo o x a coordenada em pixel no eixo horizontal, y a coordenada em pixel no eixo vertical e z a coordenada de profundidade dada em milímetros.

Uma EF pode possuir uma ou mais funções sintáticas no contexto das LS. Neste contexto defini-se nove funções sintáticas, as quais estão descritas na Tabela 6. Ao assumir uma função sintática, a EF é considerada uma EF gramatical (EFG). Neste trabalho, define-se, então, o mapeamento entre funções sintáticas e EFGs ilustrado nas primeira e segunda colunas da Tabela 6. As demais colunas dessa tabela descrevem as características físicas atemporais (configuração dos elementos da face – colunas 3, 4 e 5) e temporais (movimento da cabeça – coluna 6). A Tabela 7 descreve os caracteres que foram utilizados para representar as características físicas e os movimentos na Tabela 6. Observe que em termos de descrição via características da face, EFGs de frases interrogativas (s/n) e condicionais, assim como, de frases com tópico e foco, podem assumir as mesmas configurações de face[1].

Vale ressaltar que não é objetivo desta dissertação de mestrado determinar a tradução semântica do que está sendo sinalizado, mas sim identificar qual configuração a face assumiu durante determinado período. Assim, o reconhecimento aqui proposto assume um caráter descritivo da LS.

A estratégia para resolução do problema de reconhecimento das EFGs adotada neste trabalho foi modelada para resolução de um probema de classificação binário, onde o modelo é preparado para identificar a ocorrência de uma EFG (classe positiva) dentro de uma frase sinalizada.

Adotar uma estratégia de classificação binária, nesta dissertação, forneceu condições para a realização de um estudo sobre a complexidade do problema estudado. O reco-

[1] Para efeitos dos experimentos realizados neste trabalho, tais EFGs assumem as mesmas configurações de face.

Funções sintáticas	**Id.**	**Sobrancelha**	**Olhos**	**Boca**	**Cabeça**
Interrogativa (qu)	EF_2	$\downarrow$			$\uparrow$
Interrogativa (s/n) / Condicional	EF_3	$\uparrow$			$\downarrow$
Interrogativa (dúvida)	EF_4	$\downarrow$	$*$	$*$	$\ominus$
Negativa	EF_1	$\downarrow$		$\cap$	$\leftrightarrow$
Afirmativa	EF_5				$\updownarrow$
Relativa	EF_6	$\uparrow$			
Tópicos / Foco	EF_7	$\uparrow$	$\diamond$		$\downarrow$

Tabela 6 – Expressões Faciais Gramaticais: mapeamento considerando as funções sintáticas; descrição considerando características físicas atemporais e temporiais.

Caracter	**Descrição**
$\uparrow$	Movimento para cima
$\downarrow$	Movimento para baixo
$\leftrightarrow$	Movimento para direita e para esquerda
$\updownarrow$	Movimento para cima e para baixo
$*$	Comprimido
$\diamond$	Aberto
$\ominus$	Afastar
$\cap$	Cantos da boca para baixo

Tabela 7 – Descrição dos caracteres utilizados na Tabela 5

nhecimento automático de EFGs na Libras ainda não havia sido estudado pela área de Computação, até o momento de desenvolvimento deste trabalho. Portanto, pouco se sabia sobre a complexidade envolvida no problema. De fato, a complexidade inicial aqui atribuída ao problema foi derivada do estudo sobre o reconhecimento de EFs em outros contextos, como no caso de reconhecimento de expressões afetivas e no contexto de outras línguas de sinais.

A estratégia adotada faz uso dos dados em sua forma original e também em representação vetorial. O conjunto de dados utilizado no presente estudo, bem como os procedimentos aplicados sobre os dados para pré-processamento e construção de representação vetorial, são apresentados nas duas primeiras seções deste capítulo. Na sequência, a terceria seção apresenta o estudo de experimentação adotado. Os resultados obtidos bem como as respectivas análises estão descritos no Capítulo 6.

5.1 Conjunto de Dados

Durante as leituras realizadas na execução do levantamento bibliográfico foi observado que os diferentes trabalhos de pesquisa reportados nos artigos científicos fazem uso

de diferentes abordagens para aquisição e construção do conjunto de dados que são usados nos seus experimentos. Muitos deles fazem uso de técnicas para extração de características a partir de imagens adquiridas via câmeras filmadoras e outros fazem uso de sistemas de sensoreamento.

Neste trabalho optou-se pelo uso do sensor Microsoft Kinect^{TM}[2]. Tal escolha foi motivada pela praticidade de uso do sensor (observada durante experimentos preliminares de aquisição de dados realizados no início de desenvolvimento deste trabalho), pela qualidade dos dados adquiridos com seu uso e também por ser um sensor de baixo custo. Detalhes sobre a aquisição de dados realizada com esse dispositivo são discutidos na Seção 5.1.1.

Também foi necessário projetar e produzir um conjunto de dados próprio para execução do estudo aqui proposto. Isso porque não havia, até então, a disponibilização de dados referentes à Libras que permitisse diretamente o estudo do reconhecimento das EFGs. Embora existam vários vídeos, disponibilizados publicamente, que contêm cenas de pessoas usando a Libras em diferentes contextos, o uso de tais vídeos implicaria em complexidade de extração de dados e interpretação da língua que levariam a execução deste trabalho para além do escopo objetivado. Também foi necessário executar alguns procedimentos de pré-processamento a fim de anular variações referentes à translação e posicionamento do sinalizador em frente ao sensor. Detalhes sobre o conjunto de dados construído são apresentados na Seção 5.1.2. Os procedimentos de pré-processamento são apresentados na Seção 5.1.3.

5.1.1 Aquisição de dados

A aquisição dos dados foi realizada com o uso do *Microsoft Face Tracking Software Development Kit for Kinect for Windows* (*Face Tracking SDK*)[3] - um mecanismo especialmente projetado para suportar rastreamento de faces imageadas pelo sensor Kinect. A *Face Tracking SDK* pode ser acessada a partir de um projeto em C++, sob o uso da *IDE Microsoft Visual Studio Express Edition for C++*. A partir do uso das diferentes funções disponíveis nesse pacote de desenvolvimento, foi possível construir uma aplicação capaz de extrair pontos da face de uma pessoa posicionada em frente ao dispositivo Kinect.

[2] Dispositivo capaz de capturar imagens RGB, imagens dotadas de informações de profundidade e também gravar informação sonora (`http://msdn.microsoft.com/en-us/library/hh855347.aspx`).

[3] `http://msdn.microsoft.com/en-us/library/jj130970.aspx`

A aplicação desenvolvida é capaz de capturar em torno de 27 *frames* por segundo, em tempo real, utilizando um computador `Intel® Core(TM) i5-3317CPU 1.70GHZ, 4GB de Memória, HD SSD 32Gb`, um sensor Kinect modelo 1414 e programação *multithreading*. A aplicação disponibiliza a imagem de cada *frame* (Figura 14(a)) e os 100 pontos extraídos da face (Figura 14(c)), para cada um dos *frames*. Mais detalhes sobre os pontos obtidos podem ser verificados no Apêndice C.

Figura 14 – Exemplo de captação de pontos da face humana, realizada com o uso do Face Tracking SDK e do sensor Kinect: (a) imagem da face a partir da qual os pontos em (b) foram extraídos; (c) visão detalhada dos pontos fornecidos pela aplicação de aquisição de dados.

A biblioteca *Face Tracking SDK* utiliza duas formas de capturar os pontos da face, sendo a primeira chamada de *StartTracking* e a segunda *ContinueTracking*. A primeira função é utilizada para inicar o processo de aquisição dos dados. A segunda função, executada na sequência da primeira, utiliza informações do *frame* anterior para identificação dos pontos da face no *frame* atual. Trata-se de um processo de predição de pontos que auxilia na obtenção de informações que permitem a aquisição de uma quatidade maior de *frames* por segundo (em torno de 27 *frames* por segundo). O uso da *ContinueTracking* é opcional, no entanto, com o uso apenas da *StartTracking*, a taxa de captação de *frames* por segundo é bem menor, uma vez que ela não possui a capacidade de uso de informação apriori para apoiar a análise do *frame* atual (*frame* sendo capturado). Devido à sua limitação de uso da informação do *frame* anterior, a *StartTracking* apresenta custo computacional alto e desempenho não satisfatório para os objetivos do presente estudo, uma vez que a reconstrução dos *frames* capturados em um vídeo não representam o conteúdo com fluidez.

Embora a ferramenta de extração de dados seja capaz de fornecer vários pontos da face com uma precisão adequada para a realização do reconhecimento de padrões das EFGs, ela apresenta problemas para lidar com oclusões. Naturalmente, uma língua de sinais é composta por sinais que possuem a região da cabeça como ponto de articulação e, nesses casos, invariavelmente, as mãos do sinalizador se colocam entre a sua face e o

sensor de captura dos pontos da face. A captura oferecida pelo sensor Kinect é prejudicada quando há oclusões. O sensor juntamente com as funções da *Face Tracking SDK* são capazes de recuperar a qualidade da captura quando do fim da oclusão, no entanto, os *frames* capturados enquanto a oclusão ocorre não são bem representados.

5.1.2 Organização dos dados

Para construção do conjunto de dados, foram escolhidas cinco frases que envolvem cada uma das EFGs de interesse neste trabalho. As frases foram compostas com sinais que evitassem oclusões da face (pelas mãos), e foram captadas a partir de cinco execuções diferentes realizadas por dois sinalizadores fluentes em Libras. Essas frases são apresentadas na Tabela 8.

Tabela 8 – Conjunto de frases que compõem o contexto do conjunto de dados.

Interrogativa (qu)	**Interrogativa (s/n)**	**Interrogativa (dúvida)**
Quando a Waine pagou?	Waine comprou um carro?	Waine comprou UM CARRO?
Porque a Waine pagou?	Isso é seu?	Isso é SEU?
O que é isso?	Você se formou?	Você se FORMOU?
Como faz isso?	Você gosta de mim?	Você gosta DE MIM?
Onde você mora?	Você vai embora?	Você vai EMBORA?
Negativa	**Afirmativa**	**Condicional**
Eu não vou.	Eu vou.	Se chover, eu não vou.
Eu não fiz nada.	Eu quero.	Se você faltar, você vai perder.
Eu nunca fui preso.	Eu gosto.	Se você não quiser, ele aceita.
Eu não gosto.	Eu comprei.	Se você não comprar, ele vai querer.
Eu não tenho.	Eu trabalho lá.	Se fizer sol, eu vou pra praia.
Relativa		
Menina que caiu de biclicleta, ela está no hospital.		
A Universidade Unifei, ela fica em Itajubá.		
Aquela empresa, ela trabalha com tecnologia.		
A Waine, amiga do Lucas, é formada em pedagogia.		
A Celi, escola de surdos, fica em SP.		
Tópicos		**Foco**
Universidade, eu estudo na USP.		Foi a WAINE que fez.
Frutas, eu gosto de abacaxi.		Eu gosto de AZUL.
Trabalho, eu trabalho com informática.		A WAINE que pagou.
Computador, eu tenho um notebook.		A bicicleta QUEBROU!
Esporte, eu gosto de volei.		VOCÊ que está errado.

As frases foram executadas em sequência, cinco vezes, em uma única sessão de captação de dados. Os arquivos de dados foram armazenados separadamente por contexto sintático (cada conjunto de frases com um tipo de EFG), e contemplam as imagens

originais, as imagens originais com os pontos (x, y) plotados, imagens de fundo branco com os pontos plotados, e os respectivos vídeos. As coordenadas (x, y, z) de 100 pontos de cada *frame* estão armazenadas em arquivos do tipo texto juntamente com um índice (tempo em milissegundos) que indica qual é o arquivo de imagem correspondente.

Todos os dados foram rotulados (1 – expressão ou classe positiva e 0 – não-expressão ou classe negativa) por dois codificadores, fluentes em Libras e responsáveis também pelas sinalizações usadas na geração do conjunto de dados. A rotulação fornecida pelo codificador 1 foi usada como os rótulos dos dados no aprendizado supervisionado aplicado na implementação do reconhecimento das EFGs.

A Tabela 9 descreve o conjunto de dados construído em termos de quantidade de *frames* captados, quantidade de *frames* positivos e quantidade de *frames* negativos para cada tipo de frase e para cada sinalizador; e a quantidade total de *frames* captados. A rotulação do codificador 1 foi usada como base para descrever o conjunto de dados.

Ainda nessa mesma tabela são apresentados os coeficientes de concordância de rotulação, obtido na comparação das rotulações dos dois codificadores. A medida de concordância usada é o Krippendorff's Alpha (ARTSTEIN; POESIO, 2008). Esse coeficiente varia de -1 a 1. Valores negativos indicam que a rotulação foi aleatória ou de concordância insuficiente; valores entre 0 e 0,2 indicam uma concordância leve; entre 0,2 e 0,4, uma concordância justa; entre 0,4 e 0,6, moderada; entre 0,6 e 0,8, substancial e, finalmente, entre 0,8 e 1, uma concordância perfeita.

A apresentação desta medida pode ajudar no entendimento da dificuldade inerente ao problema. Observe que a concordância de rotulação entre dois humanos só pode ser considerada perfeita em cinco casos. No entanto, a medida também fornece robustez para a rotulação usada para treinamento dos modelos neste trabalho, uma vez que a menor concordância (que seria a mais problemática) pode ser considerada de confiança moderada/substancial.

5.1.3 Pré-processamento dos dados

Antes da realização dos experimentos, duas estratégias foram adotadas para o pré-processamento dos dados: a translação e normalização. Elas foram necessárias para redução de ruídos e da influência de outras informações que não fazem parte do problema de reconhecimento das expressões: a localização do sinalizador nos eixos x e y em relação

Tabela 9 – Descrição do conjunto de dados.

EFG	**Sinalizador 1**			**Sinalizador 2**			**Total**	**Coef. de Concordância**	
	+	-	**Total 1**	+	-	**Total 2**			%
Int. (qu)	609	677	1286	549	779	1328	2614	0,83	0,92
Int. (s/n)	532	858	1390	715	1023	1738	3128	0,67	0,86
Int. (dúvida)	491	821	1312	780	717	1497	2809	0,77	0,90
Negativa	528	596	1124	712	870	1582	2706	0,83	0,92
Afirmativa	414	648	1062	528	546	1074	2136	0,60	0,82
Condicional	548	1359	1907	589	1445	2034	3941	0,82	0,93
Relativa	644	1686	2330	550	1354	1904	4234	0,90	0,96
Tópicos	360	1436	1796	467	1358	1825	3621	0,80	0,94
Foco	330	1073	1403	531	813	1344	2747	0,81	0,94

ao posicionamento do sensor; e as variações no eixo z, que representa a distância entre o sinalizador e o sensor.

A primeira estratégia utilizada foi a translação de todos os pontos para uma mesma referência base, escolhida para representar a origem do espaço $(0, 0, 0)$. O ponto referência escolhido para translação foi a ponta do nariz. Neste procedimento, a média de todos os pontos do nariz (média em x, média y) foi subtraída dos demais pontos, nos eixos x e y. A média dos pontos do nariz de cada sinalizador foi aplicada, separadamente. Na Figura 15 é possível observar os efeitos da aplicação de tal procedimento.

(a) Antes (b) Depois

Figura 15 – Sobreposição de fotos dos dois sinalizadores, sendo o sinalizador 1 em preto e sinalizador 2 em vermelho, demonstrando a necessidade de translação das faces para um mesmo ponto em comum.

A segunda estratégia utilizada foi a normalização, cujo objetivo foi eliminar a influência da distância entre sinalizador e sensor (eixo z). Para tal, ao invés de usar os valores absolutos provenientes dos cálculos das variações de distâncias euclidianas e ângulos (veja Seção 5.2.1), optou-se por usar a variação relativa. Para normalização dos pontos, utilizou-se somente os valores máximos e mínimos de cada distância separadamente.

5.2 Representação dos Dados

Os dados brutos, como são extraídos a partir do uso do sensor, podem ser diretamente usados como características descritivas do dado e podem fazer parte, ou constituir, o vetor de características a ser usado como entrada dos modelos de classificação supervisionada. No entanto, é interessante derivar informações com potencial para descrever melhor as caracerísticas que marcam cada um dos aspectos de uma EF. Também, a quantidade de pontos capturados pelo sensor, pode trazer informação correlacionada, cujo uso não se faz adequado em um vetor de características por aumentar a complexidade do espaço de características

Assim, esta seção se destina a descrever as estratégias aplicadas para realização da extração de características a serem usadas neste trabalho (Seção 5.2.1) e os vetores para representação dos dados resultantes (Seção 5.2.2).

5.2.1 Extração de Características

O sensor Kinect combinado à aplicação de funções da *Face Tracking SDK* possibilita a captura de 100 pontos da face sob sensoreamento. Porém, muitos desses pontos possuem alta correlação, uma vez que dizem respeito à descrição de regiões bastante próximas dentro da área de movimentação dos elementos da face (boca, nariz, olhos, sobrancelhas e contorno da face). Por esse motivo, foi realizado um estudo sobre a correlação existente entre esses pontos.

Nesse estudo, uma medida de correlação entre os pontos foi calculada considerando o deslocamento sofrido por ele quando há movimentação dos elementos da face. Os movimentos considerados nesse estudo são todos aqueles que podem ocorrer nas EFGs da Libras. Foram gravados três vídeos, sobre os quais os 100 pontos da face foram capturados, considerando os movimentos possíveis dos elementos da face e cabeça, conforme especificado abaixo:

1. Movimentos da sobrancelha (levantando e contraindo), boca (aberta e fechada, comprimida e com abertura normal) e cabeça (movimento vertical e horizontal).
2. Movimentos da sobracelha (levantando e contraindo) e boca (aberta e fechada, comprimida e com abertura normal).

3. Movimentos da cabeça (aberta e fechada, comprimida e com abertura normal).

Cada vídeo compôs um experimento onde os 100 pontos foram analisados de acordo com a correlação existente entre eles. A cada iteração do experimento, a maior medida de correlação encontrada indicava pares de pontos para serem substituídos pelo seu ponto médio. Esse procedimento foi iterativamente executado enquanto houvesse uma medida de correlação maior ou igual a 0,65, partindo de uma correlação em 1 e subtraindo 0,005 de cada iteração. Contudo, substituir pares de pontos com correlação mais baixa do que o valor 0,97 levou à descaracterização da representação de uma face, por exemplo, o agrupamento dos pontos da sobrancelha com os pontos dos olhos. Portanto, a medida de correlação 0,97 foi empiricamente determinada como a correlação mínima a ser considerada para substituição de pares de pontos. O resultado da substituição de pontos obtido com esse procedimento pode ser visto na Figura 16.

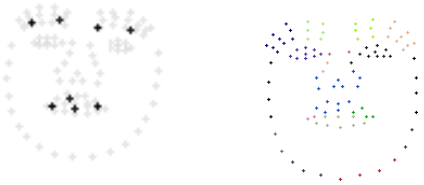
(a) Grupo 1 (b) Grupo 2 (c) Grupo 3

Figura 16 – Pontos de mesma cor são aqueles que devem ser agrupados (ou substituídos por seu ponto médio). Correlação mínima considerada: 0,97.

Como pode ser observado na figura 16(c) (resultados obtidos do experimento com o vídeo 3 - movimentos da cabeça), praticamente todos os pontos dos principais elementos da face (boca, nariz, olhos e sobrancelhas) são altamente correlacionados. Esse fato indica que dentre os pontos desses elementos, qualquer um deles poderia ser escolhido para trazer informação sobre movimento da cabeça nos eixos x e y. Portanto, fez-se interessante analisar os pontos desses elementos sem considerar a movimentação da cabeça, de forma a verificar como eles se comportam em relação à representação dos movimentos da face. Assim, pontos selecionados a partir da análise do Grupo 2 (figura 16(b)) foram escolhidos para a continuidade dos experimentos desta dissertação. Os pontos do grupo 1 não foram utilizados na análise, uma vez que se constatou que a informação do movimento da cabeça poderia ser extraído de qualquer ponto.

O conjunto de pontos final é composto por 8 pontos conforme ilustrado na figura 17(a). É importante ressaltar que o conjunto de pontos obtidos está em conformidade

com Chang e Huang (2010) e Wang et al. (2010) que utilizam os mesmos pontos em seus trabalhos; além disso, o conjunto final de pontos é também semelhante ao que foi usado nos trabalhos de Dahmane e Meunier (2012), Nguyen e Ranganath (2012) e Yu, You e Tsai (2012), que acrescentam somente alguns pontos a mais entre a boca e o nariz.



(a) Grupo 1 (b) Grupo 2

Figura 17 – Representação final dos pontos escolhidos, com a imagem de todos os pontos em marca d'água no fundo em (a) e pontos selecionados em vermelho em (b).

A partir do conjunto de 8 pontos selecionados, para cada *frame* optou-se por extrair dois conjuntos de medidas para serem usadas na composição dos vetores de características: D, com 28 medidas de distância, e A com 168 ângulos, calculadas de acordo com o tipo de vetor de características sendo formado (veja Seção 5.2.2). Sendo $D = \{D_1, D_2, ...D_{28}\}$ as distâncias entre os pontos que descrevem a face e $A = \{A_1, A_2, ...A_{168}\}$ os ângulos que são formados entre eles.

Além disso, uma representação de dados em janelas é usada em complementação à representação *frame* a *frame*, de forma a caracterizar a informação sobre a movimentação dos elementos da face no tempo. O parâmetro J define o tamanho de uma "janela" em *frames*, considerando sequência de *frames*. A Tabela 10 ilustra janelas de três tamanhos diferentes. Para o caso de janelas do tamanho $J = 1$, considera-se o problema em seu aspecto atemporal.

Tabela 10 – Exemplos de janelas de três tamanhos diferentes: 1, 2 e 3.

Tam.	**Janela 1**	**Janela 2**	...	**Janela m**
1	$\{frame_1\}$	$\{frame_2\}$	...	$\{frame_n\}$
2	$\{frame_1; frame_2\}$	$\{frame_2; frame_3\}$	...	$\{frame_{n-1}; frame_n\}$
3	$\{frame_1; frame_2; frame_3\}$	$\{frame_2; frame_3; frame_4\}$	...	$\{frame_{n-2}; frame_{n-1}; frame_n\}$

No caso de uso de uma representação com janelas de tamanho $J > 1$, faz-se necessário definir o "*frame* de interesse", i.e., o *frame* ao qual aquela representação em janela se refere, sendo que para este trabalho, utilizou-se o primeiro *frame* da janela. Em consequência desta escolha, para cada experimento com janela maior que 1, o conjunto de dados foi reduzido no montante de *frames* referente ao tamanho da janela. No entanto,

como os últimos *frames* sempre eram "não-expressão" e, devido ao tamanho das janelas (geralmente menores que 11 *frames*), não houve prejuízo significativo à análise dos resultados de reconhecimento. Em um conjunto com 1200 *frames*, para uma janela de 10 *frames*, o conjunto final ficou com 1191 *frames*, perdendo-se apenas os últimos 10 *frames*.

5.2.2 Vetores de Características

Com o intuito de explorar diferentes formas de organizar as características extraídas da face, algumas combinações dessas características são propostas para compor o vetor de características a ser usado nos experimentos desta dissertação. Seis tipos de vetores de características estão sendo usados:

- Vetor XY: Vetor com informações referentes às coordenadas de (x,y) dos pontos.
- Vetor XYZ: Vetor com informações referentes às coordenadas de (x,y,z) dos pontos.
- Vetor 1: Vetor com informações derivadas das coordenadas (x, y) de cada ponto.
- Vetor 2: Vetor com informações derivadas das coordenadas (x, y, z) de cada ponto.
- Vetor 3: Vetor com informações derivadas das coordenadas (x, y) de cada ponto, mais a informação de profundidade de cada ponto em separado, i.e., utilizando a informação z de todos os pontos de um *frame* também na composição do vetor.
- Vetor 4: Vetor com informações derivadas das coordenadas (x, y, z) de cada ponto, mais a informação de profundidade de cada ponto em separado, i.e., utilizando a informação z de todos os pontos de um *frame* também na composição do vetor.

A tabela 11 mostra as combinações de informações presente em cada vetor, atribui uma abreviação para elas e também as relaciona com os tipos de vetores referentes aos itens que foram levados em consideração, tais como distância, ângulos e pontos de referência. No total, 42 vetores de características foram criados para o presente estudo.

Para os vetores que não tiveram pontos de referência acrescentados, tem-se 28 medidas de distância, 168 medidas ângulos, 8 dados da coordenada z. Para encontrar o tamanho deste vetor, basta somar sua composição, como por exemplo o Vetor 3 com todas as distâncias e todos os ângulos que possui 204 informações (28 distâncias + 168 ângulos + 8 valores de profundidade de cada ponto). Para os vetores que utilizaram mais um ponto como referência no cálculo das distância e dos ângulos, tem-se o cálculo de todas as distâncias e ângulos com 9 pontos, totalizando 36 distâncias e 252 ângulos. Para os

Tipo de vetor	**Informação**	**Abreviação**
XY	coordenadas	XY
XYZ	coordenadas	XYZ
1, 2, 3, 4	todas as distâncias e todos os ângulos	DA
	todas as distâncias	D
	todos os ângulos	A
medidas considerando também um ponto de referência nos olhos		
1, 2, 3, 4	todas as distâncias e todos os ângulos	DAO
	todas as distâncias	DO
	todas os ângulos	AO
medidas considerando também um ponto de referência no nariz		
1, 2, 3, 4	todas as distâncias e todos os ângulos	DAN
	todas as distâncias	DN
	todas os ângulos	AN
1, 2, 3, 4	distâncias e ângulos de Yu, You e Tsai (2012) distâncias de Chang e Huang (2010) ângulos de Wang et al. (2010)	DAQ

Tabela 11 – Descrição dos vetores de características. Abreviações: XY e XYZ: coordenadas; D: distâncias; A: ângulos; O: olhos; N: nariz; Q: referência na literatura.

vetores com janelas, como utilizou-se as informações de J *frames*, multiplica-se o tamanho do vetor utilizado pelo tamanho da janela.

5.3 Experimentos: classificação binária

Os experimentos executados no âmbito desta dissertação dizem respeito à criação de uma série de classificadores binários, implementados por meio de redes neurais Percetron Multicamadas, treinadas com o algoritmo de Retropropagação do Erro. O problema de reconhecimento de EFGs é modelado considerando que:

- O dado a ser reconhecido vem da sinalização de uma frase em Libras, contendo o uso de uma EFG de interesse, gravada em um vídeo S que deve ser visto como uma sequência de *frames* $\{f_1, f_2, ..., f_n\}$;
- Uma representação vetorial V, contendo informações extraídas de cada um dos *frames* do vídeo, é usada como entrada para um modelo classificador, podendo considerar uma janela de *frames* a fim de propiciar uma representação espacial do tempo onde um movimento ocorre;

- O classificador analisa a informação referente a cada um dos *frames* de entrada (único ou em janela), a fim de decidir se ele faz parte, ou não, do trecho de vídeo no qual a EFG de interesse ocorre;
- A resposta Y do classificador $\in \{+1, -1\}$, sendo que $+1$ significa que o *frame* pertence ao trecho de vídeo no qual a EFG ocorre; e -1 significa que o *frame* não pertence ao trecho de vídeo no qual a EFG ocorre.

Foram treinadas diferentes redes neurais considerando:

- Todas as possibilidades de vetores de características apresentadas na Seção 5.2.2;
- Variações nos tamanhos de janelas, considerando o intervalo $[1, N_{max}]$, onde N_{max} corresponde à metade do número de *frames* no menor trecho de vídeo que contém a EFG de interesse[4];
- Variações nos parâmetros de treinamento da rede neural, considerando as seguintes variações:
  - algoritmo de aprendizado *backpropagation*.
  - número de neurônios na camada escondida, variando 5 para mais e para menos do resultado da raiz quadrada do tamanho do vetor de entrada.
  - taxa de aprendizado iniciando em 1, sendo dividida por 2 até o valor 0,0156.
  - em testes preliminares, identificou-se que o número de épocas não alterava o resultado, então utilizou-se o valor de 200 épocas.
- Variações no modelo de treinamento e teste no que diz respeito à sinalização das frases em Libras, compondo o seguinte conjunto de experimentos.
  - Experimento 1: treinamento com sinalizador 1 teste com sinalizador 1.
  - Experimento 2: treinamento com sinalizador 2 teste com sinalizador 2.
  - Experimento 3: treinamento com sinalizador 1 teste com sinalizador 2.
  - Experimento 4: treinamento com sinalizador 2 teste com sinalizador 1.
  - Experimento 5: treinamento com sinalizador 1 e 2 teste com sinalizador 1 e 2.

Frases completas são usadas para o treinamento, validação e testes dos classificadores. Uma vez que os *frames* devem ser apresentados ao classificador considerando a sequência

[4] Este parâmetro assume valores diferentes em cada experimento, considerando sempre o menor tempo total para execução de uma expressão nas frases e, portanto, evitando que uma "janela" seja grande o suficiente para conter *frames* que representem: `não-expressão -- expressão -- não-expressão`

em que eles compõem o vídeo, a estratégia de uso das frases pode ser considerada adequada. Testes preliminares mostram que a apresentação aleatória dos *frames*, durante as fases de treinamento, validação e teste, gerava resultados surpreendentemente bons. No entanto, tais resultados eram obtidos principalmente com o uso de janelas, mostrando que a apresentação aleatória em combinação com o uso das janelas estava, na realidade, propiciando que todo o vídeo fosse apresentado para o classificador já na fase de treinamento.

5.4 Considerações Finais

Este capítulo apresentou a forma com que os dados foram adquiridos, organizados e apresentados à rede neural, além de detalhar a combinação dos testes que foram realizados e o tratamento que essas informações receberam com objetivo de eliminar ruídos nos dados de entrada. Os resultados dos experimentos com classificadores binários para reconhecimento de EFGs são apresentados no Capítulo 6.

6 Reconhecimento de Expressões Faciais Gramaticais: resultados e análises

O presente capítulo visa apresentar os resultados finais encontrados na resolução do problema aqui proposto. Para organizar a apresentação dos resultados, este capítulo é dividido por expressões faciais, considerando análises particulares de cada expressão facial, e por fim, uma análise geral sobre as características em comum das EFGs. Vale lembrar que foram realizados cinco experimentos em cada EF: sendo os experimentos 1 e 2 realizados, cada um, com um sinalizador diferente; experimento 3 realizado com os dados do primeiro sinalizador para treino e testado nos dados do segundo sinalizador; o oposto foi feito no experimento 4, utilizando os dados do sinalizador 2 para treino e testado com os dados do sinalizador 1; e por fim, no experimento 5, dados de ambos sinalizadores foram utilizados para teste e treino.

Os resultados são apresentados de forma padronizada para cada uma das EFGs. As análises dos resultados são realizadas sob dois pontos de vista: (a) o primeiro se refere à análise do desempenho, em termos de *F-score*, dos cinco melhores classificadores; (b) o segundo diz respeito a uma análise do tipo de erro que os melhores classificadores, de cada experimento, cometeram. Ao final do capítulo é apresentado um resumo com as principais conclusões obtidas no conjunto dos experimentos.

Para possibilitar a análise (a), são apresentadas tabelas em que cada experimento os *F-scores* dos cinco melhores classificadores obtidos são apresentados considerando a representação sem janela e a representação com janelas. Para esse último caso, é apresentado o número de *frames* (J) que compõem a janela em que o resultado foi obtido. Também é mencionado na tabela o vetor de característica dos resultados obtidos.

Na análise (b) são apresentados os "erros de borda", de forma que seja possível verificar, com mais detalhes, o tipo de erro cometido pelo melhor classificador obtido em cada experimento. O "erro de borda"é definido como erros de classificação que acontencem dentro da faixa de transição entre a ocorrência da EFG sob análise e a sua "não ocorrência"(aqui chamada de fase de "não-expressão"). Neste trabalho foi arbitrada uma faixa de seis *frames* como trecho de borda (ou trecho de transição), sendo que esse trecho corresponde a três *frames* antes do início da ocorrência da expressão e três trechos após.

Como exemplo do procedimento adotado para análise de erros de borda, considere a sequência de rótulos atribuída pelo rotulador humano para uma sequência de *frames*:

00001111. Se, como resposta do modelo neural aplicado à mesma sequência de *frames*, obtem-se 01111111 ou 00000001, serão identificados três erros de borda para cada caso. Os três primeiros rótulos positivos da primeira sequência, e os três últimos rótulos negativos na segunda sequência.

As tabelas que trazem os detalhes sobre os erros de borda mostram as seguintes informações para o melhor classificador obtido em cada experimento, considerando representação sem janela e com janela[1]:

- *F-score* obtido pelo classificador[2];
- erro total cometido pelo classificador, em termos de número de *frames* classificados erroneamente e a porcentagem que ela representa do total de *frames* apresentados no teste do modelo;
- erros na borda, em termos de número de *frames* de borda classificados erroneamente e quanto isso representa do total de *frames* classificadores errados no teste;
- quantidade de *frames* de EFGs classificadas como "não-expressão";
- quantidade absoluta e relativa de *frames* de EFGs classificadas como "não-expressão" dentro do trecho de borda;
- quantidade de *frames* de "não-expressão" classificadas como EFGs;
- quantidade absoluta e relativa de *frames* de "não-expressão" classificadas como EFGs dentro do trecho de borda.

Erros de borda podem ser considerados, até um limite (arbitrário), como erros admissíveis, uma vez que o aprendizado do modelo neural está baseado em uma rotulação fornecida por um rotulador humano, ou seja, a rotulação está sujeita a imprecisões no que diz respeito à tomada de decisão sobre o *frame* exato onde uma EFG de fato se estabeleceu.

6.1 Expressão Facial Gramatical: Interrogativa (qu)

A EFG **interrogativa (qu)** é caracterizada pelo abaixamento das sobrancelhas e pelo movimento vertical da cabeça para cima. Os resultados obtidos para o reconhecimento dessa expressão são bastante uniformes (veja Tabela 12), sendo muito pequenas as diferenças

[1] O tamanho da janela é apresentado como o número subscrito que acompanha o identificador do experimento.

[2] Esse dado, assim como o tamanho da janela, são dados já apresentados na tabela da análise (a), e são repetidos aqui apenas para propiciar melhores condições ao leitor para entendimento geral dos resultados apresentados na tabela.

de desempenho entre os cinco melhores classificadores de cada experimento, entre os experimentos e entre classificadores que contam ou não com a representação janelada, não alcançando 10% de diferença.

Tabela 12 – Resultados para expressão facial gramatical **interrogativa (qu)**, em termos de vetor de características, F-score e tamanho de janelas.

Vetor	**F-score**	**Vetor**	**F-score**	**J**	**Vetor**	**F-score**	**Vetor**	**F-score**	**J**
Experimento 1					**Experimento 2**				
DAO_v3	0,8578	DAQ_v3	0,8942	6	D_v3	0,8640	DO_v3	0,8988	2
A_v1	0,8571	DO_v4	0,8903	8	DO_v3	0,8632	DAQ_v3	0,8841	4
DAQ_v3	0,8565	DO_v3	0,8826	6	DAQ_v3	0,8625	D_v3	0,8829	3
DA_v3	0,8559	D_v4	0,8817	7	DO_v1	0,8555	XYZ	0,8761	4
AN_v1	0,8559	AO_v1	0,8800	2	XYZ	0,8464	AO_v3	0,8688	2
Experimento 3					**Experimento 4**				
DO_v1	0,8320	D_v1	0,8743	2	DO_v3	0,8871	DO_v3	0,8979	3
DN_v1	0,8240	DN_v1	0,8521	3	D_v3	0,8859	D_v3	0,8961	3
D_v1	0,8215	DN_v3	0,8417	4	DN_v3	0,8828	DN_v3	0,896	3
D_v3	0,8140	DO_v3	0,8378	4	DO_v1	0,8805	DO_v1	0,8819	3
DN_v3	0,8015	D_v3	0,8333	4	D_v1	0,8782	D_v1	0,881	3
Experimento 5									
DO_v3	0,8287	DO_v3	0,8599	4					
D_v3	0,8283	D_v3	0,8541	3					
DN_v3	0,8242	DAQ_v1	0,8498	3					
D_v1	0,8182	DN_v3	0,8482	4					
XYZ	0,8177	DAQ_v3	0,8368	3					

Para o caso do reconhecimento da EFG interrogativa (qu), os vetores predominantemente presentes nos experimentos que alcançaram os melhores resultados são aqueles cujas características derivam das coordenadas (x, y), presentes na representação vetor 1 e vetor 3. Além disso, a presença da medida de profundidade como descritor das características das expressões faciais também se mostrou predominante (vetor 3). Porém, embora o movimento vertical da cabeça pudesse ser descrito com eficiência pela coordenada z, tal hipótese não se destacou, pois não há diferença significativa entre os resultados alcançados com o uso de representações do tipo vetor 1 (sem a presença da coordenada z) e do tipo vetor 3 (com a presença da coordenada z).

Os erros de borda para a EFG sob análise são expressivos (Tabela 13). Em todos os experimentos, pelo menos 30% dos *frames* classificados erroneamente ocorrem no período de transição, sendo que em alguns casos, a porcentagem de erros na borda chega a ser metade dos erros cometidos. Destes erros de transição, grande parte dos *frames* classificados como "expressão" (falso positivos), ocorrem nas transições.

Tabela 13 – Detalhes de erros de borda cometidos por classificadores no reconhecimento da EFG **interrogativa (qu)**.

Exp.	**F-score**	**Erro total**	**Erro borda**	**Falso neg.**	**Falso neg. na borda**	**Falso Pos.**	**Falso pos. na borda**
1	0,8578	62 (14%)	20 (32%)	51	9 (18%)	11	11 (100%)
1_6	0,8942	49 (11%)	14 (29%)	31	0 (0%)	18	14 (78%)
2	0,8640	45 (10%)	26 (58%)	29	15 (52%)	16	11 (69%)
2_2	0,8988	34 (7%)	13 (38%)	21	4 (19%)	13	9 (69%)
3	0,8323	181 (14%)	67 (37%)	100	45 (45%)	81	22 (27%)
3_4	0,8743	136 (10%)	57 (42%)	76	34 (45%)	60	23 (38%)
4	0,8871	142 (11%)	73 (51%)	51	15 (29%)	91	58 (64%)
4_3	0,8979	128 (10%)	55 (43%)	46	5 (11%)	82	50 (61%)
5	0,8287	193 (15%)	84 (44%)	137	44 (32%)	56	40 (71%)
5_4	0,8599	161 (12%)	59 (37%)	110	26 (24%)	51	33 (65%)

6.2 Expressão Facial Gramatical: Interrogativa (s/n)

A EFG interrogativa (s/n) é caracterizada pela suspensão das sobrancelhas e o abaixamento da cabeça, características opostas à EFG interrogativa (qu) e iguais à EFG condicional. Os resultados se mostraram similares dentro de cada experimento, com *F-scores* superiores a 90% nos experimentos 1 e 4, onde as redes neurais foram testadas com o sinalizador 1, o que pode mostrar um viés durante a sinalização ou expressões faciais com maior ênfase que o sinalizador 2. Estes resultados tiveram consonância com aqueles obtidos na EFG condicional, como esperado.

Em termos de vetor de características e das características em si, os resultados obtidos no reconhecimento da EFG interrogativa (s/n) é semelhante ao obtido para a expressão interrogativa (qu), com uma frequência maior do aparecimento das medidas de ângulos como descritor, porém ainda com presença da medida de distância. Entretanto, os resultados mostram melhoras um pouco mais fortes com o uso das janelas. Esse fato pode indicar que a execução das EFGs interrogativas (s/n) pode ser naturalmente mais intensa do que a execução da interrogativa (qu), embora ambas envolvam movimentos da cabeça e da sobrancelhas parecidos, porém contrários. Também é interessante notar que

Tabela 14 – Resultados para expressão facial gramatical **interrogativa (sn)**, em termos de vetor de características, F-score e tamanho de janelas.

Vetor	**F-score**	**Vetor**	**F-score**	**J**	**Vetor**	**F-score**	**Vetor**	**F-score**	**J**
Experimento 1					**Experimento 2**				
AN_v3	0,9179	DAN_v1	0,9412	3	DAQ_v3	0,8349	XY	0,9129	6
AN_v1	0,9153	DAN_v3	0,9406	4	AN_v1	0,8214	XYZ	0,9095	6
A_v3	0,9148	DO_v1	0,9398	3	A_v1	0,8172	DAQ_v3	0,8904	4
DAN_v3	0,9143	DAO_v3	0,9395	3	DA_v3	0,8161	AN_v3	0,8753	5
DN_v1	0,9135	A_v1	0,9393	3	DAO_v3	0,8157	AN_v1	0,8732	6
Experimento 3					**Experimento 4**				
AN_v1	0,7788	DAN_v3	0,8365	6	A_v1	0,9132	A_v3	0,9445	4
DAN_v1	0,7712	DAQ_v1	0,832	5	DAN_v1	0,9128	A_v1	0,9382	3
A_v3	0,7696	DAO_v1	0,8231	5	AN_v3	0,9109	AO_v1	0,9373	6
DA_v3	0,769	DA_v1	0,8201	6	A_v3	0,9089	DAO_v3	0,9363	4
A_v1	0,7676	DAN_v1	0,82	5	DAN_v3	0,9088	DAN_v3	0,9363	3
Experimento 5									
AO_v3	0,8341	XY	0,886	5					
AO_v1	0,832	XYZ	0,8809	6					
DAO_v1	0,8298	DA_v1	0,8712	6					
XY	0,8289	AO_v3	0,8668	3					
DAO_v3	0,8246	AO_v1	0,865	5					

o experimento 4 apresenta resultados superiores a 87%, mesmo sendo um experimento onde o treino é realizado com os gestos do sinalizador 2 e o teste é feito sobre os gestos do sinalizador 1. Para esses classificadores, reconhecer o padrão de marcação do sinalizador 1, mais intenso, é uma tarefa mais fácil mesmo quando o treinamento se dá com uma sinalização de padrões mais sutis (sinalizador 2).

Os resultados para essa expressão gramatical possuem uma melhoria quando os erros de borda são considerados. O número de erros que ocorrem nas bordas da execução de uma expressão são significativamente altos na maioria dos experimentos, chegando a representar 67% dos erros cometidos por um classificador, como pode ser visto na Tabela 15.

Tabela 15 – Detalhes de erros de borda cometidos por classificadores no reconhecimento da EFG **interrogativa (sn)**.

Exp.	**F-score**	**Erro total**	**Erro borda**	**Falso neg.**	**Falso neg. na borda**	**Falso Pos.**	**Falso pos. na borda**
1	0,9179	34 (7%)	19 (56%)	16	12 (75%)	18	7 (39%)
1_3	0,9412	25 (5%)	6 (24%)	6	0 (0%)	19	6 (32%)
2	0,8349	72 (12%)	38 (53%)	48	23 (48%)	24	15 (63%)
2_6	0,9129	42 (7%)	28 (67%)	10	5 (50%)	32	23 (72%)
3	0,7788	326 (19%)	140 (43%)	141	72 (51%)	185	68 (37%)
3_6	0,8365	235 (14%)	89 (38%)	114	31 (27%)	121	58 (48%)
4	0,9132	92 (7%)	58 (63%)	48	32 (67%)	44	26 (59%)
4_4	0,9445	59 (4%)	26 (44%)	30	9 (30%)	29	17 (59%)
5	0,8341	206 (13%)	110 (53%)	108	61 (56%)	98	49 (50%)
5_5	0,8860	147 (9%)	82 (56%)	55	22 (40%)	92	60 (65%)

6.3 Expressão Facial Gramatical: Interrogativa (dúvida)

As interrogativas do tipo dúvida são caracterizadas pelo abaixamento das sobrancelhas, afastamento da cabeça e compressão da boca e dos olhos. É uma EFG com variação em vários elementos da face e também caracterizada pelo deslocamento da cabeça no eixo da coordenada z.

No caso dos classificadores construídos para essa EFG percebe-se que a frequência de resultados bons obtidos com vetores de representação envolvendo a coordenada z não foi expressiva, pois o vetor 1 (x, y) também se destacou como uma boa representação. No entanto, um número maior de *frames* na janela (11 *frames* – veja na seção 5.3) ocorreu no experimento 2; e os vetores envolvendo somente os pontos da face (x, y e x, y, z) tiveram destaque também. O uso de janelas elevou o desempenho dos modelos em todos os experimentos, totalizando uma melhora de 3% nos *F-scores*, na média.

Para essa expressão facial, percebe-se que a maior parte dos erros aconteceram nas transições, com a quantidade relativa de erros de borda superiores a 60% na maioria dos experimentos (Tabela 17).

Tabela 16 – Resultados para expressão facial gramatical **interrogativa (dúvida)**, em termos de vetor de características, F-score e tamanho de janelas.

Vetor	**F-score**	**Vetor**	**F-score**	**J**	**Vetor**	**F-score**	**Vetor**	**F-score**	**J**
Experimento 1					**Experimento 2**				
DAQ_v1	0,9461	XYZ	0,9607	4	XY	0,9432	AO_v3	0,9700	11
DA_v3	0,9366	DA_v3	0,9605	3	AO_v3	0,9251	XY	0,9676	10
DAQ_v3	0,9362	DAQ_v3	0,9605	3	DAQ_v1	0,9244	DAO_v1	0,9658	11
DAN_v1	0,9337	XY	0,9602	3	DAQ_v3	0,9241	DAQ_v3	0,9657	7
DA_v1	0,9309	DAN_v1	0,9598	3	AO_v1	0,9231	AO_v1	0,9636	13
Experimento 3					**Experimento 4**				
DAQ_v3	0,8391	XYZ	0,9052	5	DO_v1	0,8933	DO_v1	0,9228	3
AO_v2	0,8254	AO_v2	0,8713	6	DO_v3	0,8929	DO_v3	0,9106	2
AO_v4	0,8239	AO_v4	0,8710	4	DAN_v1	0,8901	DAO_v1	0,9007	2
XY	0,8210	DAQ_v3	0,8657	2	DAO_v3	0,8866	XYZ	0,8965	6
XYZ	0,8179	XY	0,8655	5	DA_v3	0,8779	AO_v1	0,8940	3
Experimento 5									
XY	0,9169	XY	0,9452	5					
AO_v1	0,8976	XYZ	0,9343	5					
DAO_v3	0,8974	DAO_v1	0,9204	2					
DAQ_v1	0,8960	AO_v3	0,9161	3					
DO_v1	0,8934	AO_v1	0,9148	2					

Tabela 17 – Detalhes de erros de borda cometidos por classificadores no reconhecimento da EFG **interrogativa (dúvida)**.

Exp.	**F-score**	**Erro total**	**Erro borda**	**Falso neg.**	**Falso neg. na borda**	**Falso Pos.**	**Falso pos. na borda**
1	0,9461	18 (4%)	17 (94%)	5	4 (80%)	13	13 (100%)
1_4	0,9607	13 (3%)	12 (92%)	4	4 (100%)	9	8 (89%)
2	0,9494	24 (5%)	18 (75%)	5	5 (100%)	19	13 (68%)
2_{11}	0,9700	14 (3%)	9 (64%)	4	1 (25%)	10	8 (80%)
3	0,8391	290 (19%)	82 (28%)	24	19 (79%)	266	63 (24%)
3_5	0,9052	146 (10%)	84 (58%)	83	43 (52%)	63	41 (65%)
4	0,8933	101 (8%)	75 (74%)	68	45 (66%)	33	30 (91%)
4_3	0,9228	74 (6%)	50 (68%)	49	26 (53%)	25	24 (96%)
5	0,9169	99 (7%)	80 (81%)	37	30 (81%)	62	50 (81%)
5_5	0,9436	66 (5%)	44 (67%)	31	16 (52%)	35	28 (80%)

6.4 Expressão Facial Gramatical: Negativa

A EFG **negativa** é caracterizada pelo movimento horizontal da cabeça (direita e esquerda) e/ou movimento das sobrancelhas para baixo e/ou configuração da boca em ⋂. É uma expressão complexa, dado que diferentes possibilidades são admitidas na sua execução. Para os experimentos nessa dissertação, foram utilizadas todas as variações durante a sinalização dessa expressão, não havendo restrição durante a sinalização. No entanto, foi observado que o sinalizador 1 utiliza as configurações da boca em ⋂ em praticamente todas expressões. A Tabela 18 mostra os resultados dos cinco experimentos envolvendo esta expressão.

Tabela 18 – Resultados para expressão facial gramatical **negativa**, em termos de vetor de características, F-score e tamanho de janelas.

Vetor	**F-score**	**Vetor**	**F-score**	**J**	**Vetor**	**F-score**	**Vetor**	**F-score**	**J**
Experimento 1					**Experimento 2**				
A_v1	0,9333	A_v3	0,9582	3	A_v1	0,6868	XYZ	0,7269	6
AN_v3	0,9317	DA_v3	0,9531	3	DAQ_v3	0,6867	XY	0,7179	7
A_v3	0,9314	DA_v1	0,9510	2	DAN_v3	0,6802	DAQ_v3	0,6995	5
DAO_v3	0,9314	DAO_v1	0,9484	2	DAO_v1	0,6765	DAN_v3	0,6931	4
AO_v1	0,9307	DAO_v3	0,9484	2	DAO_v3	0,6764	DAQ_v1	0,6927	7
Experimento 3					**Experimento 4**				
DAQ_v1	0,6863	DAQ_v1	0,6760	6	DAO_v3	0,8498	D_v1	0,8806	5
DAQ_v3	0,6581	DAQ_v3	0,6617	6	DAO_v1	0,8467	D_v3	0,8691	5
AN_v1	0,6405	XY	0,6455	6	AN_v1	0,8436	DO_v1	0,8561	4
AO_v1	0,6375	DO_v4	0,6398	4	DN_v3	0,8386	A_v1	0,8541	4
DN_v4	0,6362	DN_v4	0,6396	2	DAN_v1	0,8281	AN_v1	0,8536	4
Experimento 5									
A_v3	0,7602	A_v3	0,7830	4					
AO_v1	0,7579	AO_v3	0,7673	2					
A_v1	0,7531	A_v1	0,7569	3					
AO_v3	0,7465	AO_v1	0,7478	2					
AN_v3	0,7375	AN_v1	0,7414	2					

Dentre todas as EFGs analisadas, a EFG negativa foi para a qual os resultados dos classificadores foram menos uniformes entre os diferentes experimentos. Enquanto no experimento 1 os resultados foram muito bons, nos experimentos 2 e 3 os resultados foram demasiadamente fracos.

Faz-se interessante notar, no entanto, que o experimento 4 apresentou resultados satisfatórios. Esses resultados são obtidos com classificadores que são treinados com o sinalizador 2 e testados com o sinalizador 1. Em um primeiro momento, tais resultados causam surpresa, uma vez que são melhores que os resultados obtidos com o experimento

3, onde treino e teste são compostos também por sinais de sinalizadores diferentes (o sinalizador 1 para treino e sinalizador 2 para testes). Contudo, é possível inferir a partir desse contexto, que o classificador tem certa facilidade em reconhecer a sinalização do primeiro sinalizador, que é uma sinalização mais marcada (com expressões mais bem definidas). O treinamento com expressões mais sutis (do sinalizador 2) está possibilitando, portanto, que o classificador generalize para as expressões onde os movimentos são mais intensos.

Na Figura 18 são ilustradas as faces de ambos sinalizadores em momentos de execução das expressões negativas, considerando *frames* distantes de momentos de transição, i.e., em momentos centrais da ocorrência da expressão gramatical. Note que o sinalizador 1 caracteriza de forma mais forte as variações nos elementos de sua face.

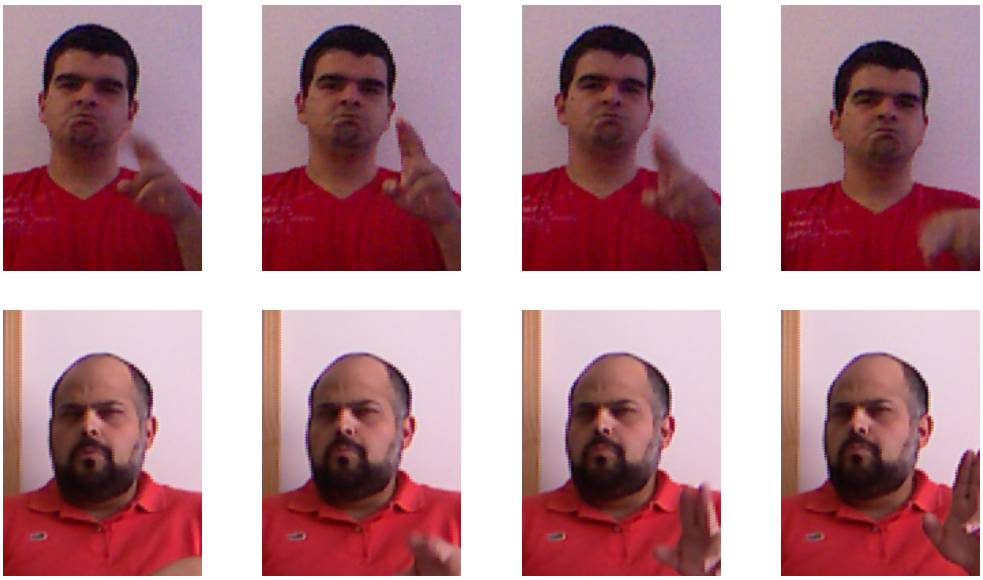


Figura 18 – Expressão Facial Gramatical Negativa: comparação entre sinalizações.

Em termos de vetores de característica e características em si, a maioria dos resultados foi obtida com vetores cujas características são derivadas das coordenadas (x, y). A presença de característica relacionadas a ângulos é bastante evidente, principalmente no experimento 5, no qual nenhum dos melhores resultados vem de classificadores construídos a partir de representações com uso de medidas de distância, por exemplo.

A análise de detalhamento dos erros de borda cometidos pelos classificadores indica que, embora haja erros na transição entre as expressões, eles não representam uma grande parcela dos erros cometidos (veja os números na Tabela 19). Isso indica que existe uma complexidade inerente à análise dessa expressão facial que se manifesta em regiões internas da ocorrência ou não da expressão.

Tabela 19 – Detalhes de erros de borda cometidos por classificadores no reconhecimento da EFG **negativa**.

Exp.	**F-score**	**Erro total**	**Erro borda**	**Falso neg.**	**Falso neg. na borda**	**Falso Pos.**	**Falso pos. na borda**
1	0,9333	27 (7%)	13 (48%)	13	8 (62%)	14	5 (36%)
1_3	0,9582	17 (4%)	7 (41%)	7	5 (71%)	10	2 (20%)
2	0,6868	166 (30%)	38 (23%)	15	3 (20%)	151	35 (23%)
2_6	0,7269	118 (21%)	25 (21%)	40	6 (15%)	78	19 (24%)
3	0,6863	640 (40%)	148 (23%)	12	1 (8%)	628	147 (23%)
3_6	0,6760	650 (41%)	141 (22%)	34	3 (9%)	616	138 (22%)
4	0,8498	152 (14%)	47 (31%)	98	25 (26%)	54	22 (41%)
4_5	0,8806	131 (12%)	55 (42%)	45	8 (18%)	86	47 (55%)
5	0,7602	330 (24%)	79 (24%)	87	25 (29%)	243	54 (22%)
5_4	0,7830	271 (20%)	68 (25%)	121	31 (26%)	150	37 (25%)

6.5 Expressão Facial Gramatical: Afirmativa

A EFG **afirmativa** é caracterizada pelo movimento vertical da cabeça (para cima e para baixo) durante a sinalização. No entanto, apesar da literatura descrever somente o movimento vertical, visualmente, percebeu-se o levantamento da sobrancelha em muitas das frases dos sinalizadores. A Tabela 20 mostra os resultados dos cinco experimentos envolvendo esta expressão.

Os vetores de características que apresentaram os melhores resultados de reconhecimento da EFG afirmativa são, em sua grande maioria, formados por características derivadas do uso apenas das coordenadas (x, y). A informação no eixo z não teve, necessariamente, influência positiva na representação do movimento vertical da cabeça, como era inicialmente esperado. Em termos de características, fica claro que distâncias e ângulos, assim como o uso de pontos de referência no cálculo das mesmas, são adequadas para descrever a expressão facial. Com relação ao uso de janelas, esses resultados mostram que o uso de janelas de *frames* traz uma melhora no desempenho dos classificadores, mostrando inclusive que os resultados com o uso de janelas são mais uniformes. Uma exceção ocorre

Tabela 20 – Resultados para expressão facial gramatical **afirmativa**, em termos de vetor de características, F-score e tamanho de janelas.

Vetor	**F-score**	**Vetor**	**F-score**	**J**	**Vetor**	**F-score**	**Vetor**	**F-score**	**J**
Experimento 1					**Experimento 2**				
DN_v1	0,8409	DA_v3	0,8773	4	DAQ_v1	0,8333	XY	0,8641	6
DAQ_v1	0,8263	AN_v1	0,8773	5	DAO_v1	0,8298	AO_v3	0,8550	4
AO_v1	0,8198	DAO_v4	0,8759	5	DAO_v3	0,8295	DN_v1	0,8522	8
DO_v3	0,8160	DO_v4	0,8759	5	AO_v3	0,8262	DAO_v3	0,8521	2
DN_v3	0,8151	DAO_v1	0,8736	5	DAQ_v3	0,8231	DAQ_v1	0,8462	3
Experimento 3					**Experimento 4**				
AO_v1	0,7469	AO_v1	0,7478	2	D_v1	0,7945	D_v3	0,8331	4
DAO_v3	0,7467	AO_v3	0,7438	2	DO_v3	0,7879	DN_v1	0,8208	5
DAO_v1	0,7443	DAQ_v3	0,7398	5	AN_v1	0,7818	DO_v1	0,8202	4
AN_v1	0,7439	AN_v1	0,7386	2	DN_v1	0,7749	DN_v3	0,8172	3
AO_v3	0,7359	DAO_v3	0,7325	2	DAN_v1	0,7746	A_v1	0,8061	4
Experimento 5									
DAN_v1	0,8057	DA_v1	0,8209	3					
DAO_v3	0,8026	AO_v1	0,8161	3					
DAN_v3	0,7995	A_v3	0,8143	3					
AN_v1	0,7987	AN_v3	0,8134	3					
AN_v3	0,7987	AO_v3	0,8133	3					

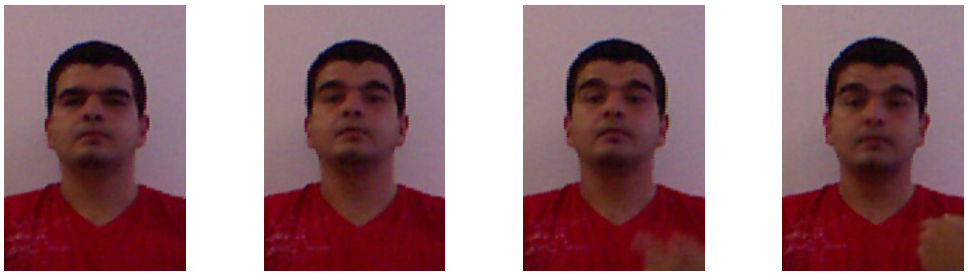
no experimento 3, onde os melhores classificadores apresentam os resultados mais baixos dentre todos os experimentos e o uso de janelas não resultou em melhorias no *F-score*.

Detalhes sobre os erros de borda cometidos pelos melhores classificadores no reconhecimento da EFG afirmativa podem ser observados na Tabela 21.

Tabela 21 – Detalhes de erros de borda cometidos por classificadores no reconhecimento da EFG **afirmativa**.

Exp.	**F-score**	**Erro total**	**Erro borda**	**Falso neg.**	**Falso neg. na borda**	**Falso Pos.**	**Falso pos. na borda**
1	0,8409	42 (11%)	20 (48%)	28	13 (46%)	14	7 (50%)
1_4	0,8773	33 (9%)	11 (33%)	21	3 (14%)	12	8 (67%)
2	0,8333	52 (14%)	22 (42%)	12	3 (25%)	40	19 (48%)
2_6	0,8641	39 (11%)	10 (26%)	18	5 (28%)	21	5 (24%)
3	0,7469	265 (25%)	92 (35%)	137	30 (22%)	128	62 (48%)
3_2	0,7478	284 (26%)	102 (36%)	107	18 (17%)	177	84 (47%)
4	0,7945	150 (14%)	75 (50%)	124	60 (48%)	26	15 (58%)
4_4	0,8331	131 (12%)	55 (42%)	87	33 (38%)	44	22 (50%)
5	0,8057	176 (16%)	78 (44%)	52	23 (44%)	124	55 (44%)
5_3	0,8227	156 (15%)	53 (34%)	55	10 (18%)	101	43 (43%)

Os erros de borda para a EFG afirmativa são mais expressivos nos experimento onde ambos os sinalizadores estão envolvidos. No experimento 4 há a maior incidência relativa dos erros comentidos nas transições. A Figura 19 ilustra uma sequência de *frames* usadas no experimento 4, a rotulação fornecida pelo especialista humano e a rotulação fornecida pelo classificador. Observe que, neste caso, houve três erros de bordas do tipo falso negativo, mas que gradativamente a rede estava caminhando para a identificação da rotulação correta.



(a) Grupo 1 (b) Grupo 2 (c) Grupo 3 (d) Grupo 3

Figura 19 – Expressão Facial Gramatical Afirmativa: análise de rotulações.

6.6 Expressão Facial Gramatical: Condicional

A expressão facial do tipo **condicional** possui as mesmas características das expressões faciais do tipo Interrogativa s/n. Ao analisar os resultados apresentados na Tabela 22, observa-se que os classifcadores apresentaram um comportamento parecido nos mesmos experimentos, demonstrando coerência entre os resultados.

Tabela 22 – Resultados para expressão facial gramatical **condicional**, em termos de vetor de características, F-score e tamanho de janelas.

Vetor	**F-score**	**Vetor**	**F-score**	**J**	**Vetor**	**F-score**	**Vetor**	**F-score**	**J**
Experimento 1					**Experimento 2**				
DN_v1	0,9459	DN_v3	0,9534	3	XY	0,7930	XYZ	0,8814	6
DN_v3	0,9434	D_v3	0,9514	2	DN_v1	0,7907	XY	0,8735	7
DO_v3	0,9412	DO_v3	0,9511	3	DO_v1	0,7897	DN_v3	0,8732	7
AO_v1	0,9396	AO_v1	0,9508	2	DO_v3	0,7778	D_v3	0,8713	7
D_v3	0,9387	DN_v1	0,9508	4	DN_v3	0,7634	DO_v3	0,8536	6
Experimento 3					**Experimento 4**				
DAN_v1	0,7530	DAN_v1	0,7704	5	DN_v3	0,9271	DAN_v1	0,9410	2
DAN_v3	0,7425	DAO_v1	0,7574	2	DN_v1	0,9261	DN_v3	0,9406	2
DAO_v1	0,7419	DAN_v3	0,7548	2	DAN_v1	0,9248	XY	0,9343	5
DAO_v3	0,7376	DAO_v3	0,7511	2	DA_v3	0,9151	DN_v1	0,9321	3
DN_v1	0,7334	DAO_v2	0,7426	6	DO_v1	0,9120	D_v1	0,9317	3
Experimento 5									
XY	0,8357	XY	0,8776	2					
AO_v3	0,8174	AN_v3	0,8319	2					
AO_v1	0,8102	AO_v3	0,8274	2					
AN_v1	0,8072	XYZ	0,8246	2					
AN_v3	0,8062	DAN_v3	0,8211	2					

Percebe-se que no experimento 2, realizado com os dados do sinalizador 2, os resultados foram bem inferiores ao experimento 1, realizado com os dados do sinalizador 1. Esta constatação reforça a ideia de expressões do sinalizador 1 são mais bem definidas, e portanto, mais fáceis de serem reconhecidas pelos classificadores.

A Tabela 23 apresenta alguns números superiores a 70% em relação aos erros de borda. Os experimentos 1 e 4 apresentam uma quantidade relativa de erros na borda altas: 80% e 71% respectivamente.

Tabela 23 – Detalhes de erros de borda cometidos por classificadores no reconhecimento da EFG **condicional**.

Exp.	**F-score**	**Erro total**	**Erro borda**	**Falso neg.**	**Falso neg. na borda**	**Falso Pos.**	**Falso pos. na borda**
1	0,9459	20 (3%)	16 (80%)	7	6 (86%)	13	10 (77%)
1_3	0,9534	17 (3%)	10 (59%)	8	3 (38%)	9	7 (78%)
2	0,7930	83 (12%)	34 (41%)	45	19 (42%)	38	15 (39%)
2_6	0,8814	49 (7%)	30 (61%)	22	14 (64%)	27	16 (59%)
3	0,7530	307 (15%)	137 (45%)	121	67 (55%)	186	70 (38%)
3_5	0,7704	292 (14%)	114 (39%)	99	49 (49%)	193	65 (34%)
4	0,9271	83 (4%)	59 (71%)	20	16 (80%)	63	43 (68%)
4_2	0,9410	66 (3%)	34 (52%)	22	4 (18%)	44	30 (68%)
5	0,8357	186 (9%)	91 (49%)	102	48 (47%)	84	43 (51%)
5_2	0,8776	140 (7%)	69 (49%)	73	37 (51%)	67	32 (48%)

6.7 Expressão Facial Gramatical: Relativa

A EFG do tipo **relativa** é caracterizada pelo levantamento das sobrancelhas e geralmente é utilizada em sentenças mais longas. Resultados superiores a 94% foram obtidos com os classificadores dedicados ao reconhecimento desta expressão (Tabela 24). Especificamente em relação à representação vetorial, os vetores do tipo vetor 1 e vetor 3 predominaram no conjunto de bons resultados, assim como a presença da informação distância nos descritores. Para essa EFG, o uso de representação janelada não trouxe resultados muito melhores do que aqueles obtidos sem o uso de janelas.

Tabela 24 – Resultados para expressão facial gramatical **relativa**, em termos de vetor de características, F-score e tamanho de janelas.

Vetor	**F-score**	**Vetor**	**F-score**	**J**	**Vetor**	**F-score**	**Vetor**	**F-score**	**J**
Experimento 1					**Experimento 2**				
DAQ_v1	0,9661	DAQ_v3	0,9680	2	DAQ_v1	0,9710	DAQ_v1	0,9759	2
DN_v3	0,9621	D_v3	0,9658	2	DO_v3	0,9679	DO_v1	0,9705	3
DO_v1	0,9619	DO_v3	0,9658	2	DO_v1	0,9650	DAO_v3	0,9702	3
DAO_v1	0,9616	DAO_v1	0,9622	2	DAO_v1	0,9544	DAQ_v3	0,9679	4
DAO_v3	0,9616	DN_v3	0,9621	3	DAO_v3	0,9511	DAO_v1	0,9674	3
Experimento 3					**Experimento 4**				
DAQ_v1	0,8717	DAQ_v1	0,8653	2	AN_v1	0,9463	AN_v1	0,9579	3
DN_v3	0,8085	DAO_v3	0,8061	3	DAN_v1	0,9417	DAN_v3	0,9549	4
DN_v1	0,8003	DAQ_v3	0,7884	2	AN_v3	0,9338	DAN_v1	0,9538	4
DAN_v1	0,7934	DAO_v1	0,7862	3	DAN_v3	0,9333	AO_v1	0,9499	3
DAO_v3	0,793	DN_v3	0,7762	2	AO_v1	0,9282	AN_v3	0,9488	5
Experimento 5									
AO_v3	0,8792	DN_v1	0,8973	3					
DAQ_v1	0,8765	DAO_v3	0,8826	4					
AO_v1	0,8736	DAQ_v1	0,8812	2					
DAO_v3	0,8731	DAO_v1	0,8790	5					
DAO_v1	0,8696	DAN_v1	0,8751	3					

Tabela 25 – Detalhes de erros de borda cometidos por classificadores no reconhecimento da EFG **relativa**.

Exp.	**F-score**	**Erro total**	**Erro borda**	**Falso neg.**	**Falso neg. na borda**	**Falso Pos.**	**Falso pos. na borda**
1	0,9661	17 (2%)	11 (65%)	6	6 (100%)	11	5 (45%)
1_2	0,9680	16 (2%)	8 (50%)	6	5 (83%)	10	3 (30%)
2	0,9710	11 (2%)	8 (73%)	3	2 (67%)	8	6 (75%)
2_2	0,9759	9 (1%)	5 (56%)	5	1 (20%)	4	4 (100%)
3	0,8717	146 (8%)	58 (40%)	54	27 (50%)	92	31 (34%)
3_2	0,8653	161 (8%)	55 (34%)	33	15 (45%)	128	40 (31%)
4	0,9463	70 (3%)	54 (77%)	27	26 (96%)	43	28 (65%)
4_3	0,9579	55 (2%)	29 (53%)	18	9 (50%)	37	20 (54%)
5	0,8792	141 (7%)	56 (40%)	75	36 (48%)	66	20 (30%)
5_3	0,8973	127 (6%)	36 (28%)	33	11 (33%)	94	25 (27%)

Devido aos resultados superiores a 86% de reconhecimento, poucos *frames* foram rotulados de maneira errônea e grande parte destes *frames* estavam presentes nos momentos de transição, como pode ser observado na Tabela 25 .

6.8 Expressão Facial Gramatical: Tópicos

A EFG do tipo **tópico** possui as mesmas características da EF do tipo **Foco**, com o levantamento das sobrancelhas, abertura dos olhos e abaixamento da cabeça. No entanto, vale ressaltar que a execução de cada tipo de EFG se diferenciam em termos de "intensidade"com que são executadas. As expressões do tipo Foco são executadas com uma intesidade maior. Os resultados apresentados pelos classificadores para ambas EFGs são similares em termos de vetores de representação mais adequados e alguma melhoria imposta pelo uso de janelas (Tabela 26 e Tabela 28), sendo esse melhoria um pouco mais evidente na EFG do tipo tópico. A questão da intensidade imposta à execução da EFG do tipo foco fica evidenciada na melhoria de resultados obtida em todos os experimentos executados com essa expressão, quando comparados com os experimentos aplicados para a EFG do tipo tópico.

Tabela 26 – Resultados para expressão facial gramatical **tópico**, em termos de vetor de características, F-score e tamanho de janelas.

Experimento 1					**Experimento 2**				
Vetor	F-score	Vetor	F-score	jan	Vetor	F-score	Vetor	F-score	Jan
DAQ_v3	0,9198	DN_v1	0,9544	5	DAQ_v1	0,8793	DAQ_v1	0,9322	3
DAQ_v1	0,9106	DO_v1	0,9426	5	DO_v1	0,875	DO_v1	0,9293	4
DO_v3	0,8971	D_v1	0,9398	6	DO_v3	0,8483	D_v1	0,9286	4
DO_v1	0,8934	D_v3	0,9398	6	DAN_v3	0,8436	DN_v1	0,9278	4
DAO_v3	0,8916	DN_v3	0,9398	5	DN_v3	0,843	DN_v3	0,9264	5
Experimento 3					**Experimento 4**				
DO_v1	0,8276	DN_v3	0,8953	5	DAO_v3	0,8780	DN_v3	0,9233	4
D_v3	0,825	D_v3	0,8874	4	DAO_v1	0,8773	DAO_v3	0,9224	3
DN_v1	0,8242	D_v1	0,8867	3	DO_v1	0,8757	DN_v1	0,9207	3
D_v1	0,823	DN_v1	0,8837	5	DO_v3	0,8729	DAO_v1	0,9205	2
DO_v3	0,8202	DA_v3	0,8801	6	AO_v1	0,8671	DO_v3	0,9201	3
Experimento 5									
DO_v3	0,8351	DO_v3	0,9164	6					
DO_v1	0,8344	DAO_v1	0,9119	5					
DAQ_v1	0,8280	DN_v3	0,9084	5					
DAN_v3	0,8231	DN_v1	0,9078	6					
DN_v3	0,8194	DO_v1	0,9049	5					

Para estas expressões, o uso de janelas teve um destaque maior que nos outros experimentos, com melhorias em todos os experimentos, com destaque para o experimento 5, que mistura os dados dos dois sinalizadores em seu treinamento. O uso de distâncias também ficou evidente nos vetores descritores, não aparecendo em somente um experimento dos 50 melhores vetores.

Os erros ocorridos nas transições são também bastante significativos para essa EFG, como mostrado na Tabela 27. Para o experimento onde houve a menor taxa de erros na borda, tal taxa ficou em 55%.

Tabela 27 – Detalhes de erros de borda cometidos por classificadores no reconhecimento da EFG **tópico**.

Exp.	**F-score**	**Erro total**	**Erro borda**	**Falso neg.**	**Falso neg. na borda**	**Falso Pos.**	**Falso pos. na borda**
1	0,9198	19 (3%)	14 (74%)	11	10 (91%)	8	4 (50%)
1_5	0,9544	11 (2%)	6 (55%)	5	2 (40%)	6	4 (67%)
2	0,8793	42 (7%)	34 (81%)	31	24 (77%)	11	10 (91%)
2_3	0,9322	24 (4%)	17 (71%)	19	12 (63%)	5	5 (100%)
3	0,8276	155 (8%)	120 (77%)	95	67 (71%)	60	53 (88%)
3_5	0,8953	95 (5%)	60 (63%)	61	31 (51%)	34	29 (85%)
4	0,8780	87 (5%)	62 (71%)	47	33 (70%)	40	29 (73%)
4_4	0,9233	55 (3%)	24 (44%)	29	8 (28%)	26	16 (62%)
5	0,8351	137 (8%)	100 (73%)	81	59 (73%)	56	41 (73%)
5_6	0,9164	71 (4%)	42 (59%)	39	20 (51%)	32	22 (69%)

6.9 Expressão Facial Gramatical: Foco

A EFG do tipo **foco** é caracterizada pelos olhos arregalados, abaixamento da cabeça e movimento para cima das sobrancelhas. Devido a semelhança das características da EFG do tipo tópico, percebe-se novamente o uso da informação distância nos vetores descritores em quase todos os experimentos. Entretanto, não houve uma melhoria com o uso de janelas na mesma proporção, apesar de resultados superiores aos resultados do tipo tópico.

Outra informação que merece destaque é o tamanho da janela que praticamente se fixou em 2. A execução de uma expressão facial do tipo foco, além de possuir uma

intensidade inerente ao contexto que se deseja destacar, ela possui como característica o seu uso somente durante a execução de um sinal, e não de uma sequência de sinais, como pode ser obsevado nas demais expressões. Essa característica pode ter sido responsável pelo tamanho da janela fixado em 2 em praticamente todos os experimentos.

Tabela 28 – Resultados para expressão facial gramatical **foco**, em termos de vetor de características, F-score e tamanho de janelas.

Vetor	**F-score**	**Vetor**	**F-score**	**J**	**Vetor**	**F-score**	**Vetor**	**F-score**	**J**
Experimento 1					**Experimento 2**				
DAO_v3	0,9630	DO_v1	0,9836	6	DAQ_v3	0,8952	DAQ_v1	0,9213	2
AO_v3	0,9627	DAO_v3	0,9835	2	DAQ_v1	0,8926	DO_v1	0,9178	2
AO_v1	0,9590	AO_v1	0,9835	2	XY	0,887	DN_v1	0,9153	2
DAO_v1	0,9587	AO_v3	0,9835	2	DO_v3	0,8851	DN_v3	0,9138	2
DAQ_v1	0,9536	DA_v1	0,9794	2	DN_v3	0,8832	DO_v3	0,9127	2
Experimento 3					**Experimento 4**				
DAQ_v1	0,8857	DAQ_v1	0,9022	2	DAO_v1	0,9120	DAO_v3	0,9538	2
DN_v1	0,8826	DO_v1	0,8984	2	DAO_v3	0,9067	DAO_v1	0,9488	2
DO_v1	0,8768	DN_v1	0,8926	2	AO_v1	0,8925	D_v1	0,9466	2
DAQ_v3	0,8584	DO_v3	0,8918	2	DO_v1	0,8901	DO_v1	0,9443	2
DO_v3	0,8434	DAO_v3	0,8745	2	DAQ_v1	0,8819	DN_v1	0,9439	2
Experimento 5									
DO_v1	0,8876	DN_v1	0,9321	2					
DO_v3	0,8861	DO_v1	0,9306	2					
DN_v1	0,8861	DO_v3	0,9266	2					
DAQ_v1	0,8852	DAQ_v1	0,9255	2					
DAO_v1	0,8831	DAO_v3	0,9180	2					

Através dos erros de transição, percebe-se que rede neural teve dificuldade em rotular os *frames* com expressão nas bordas dos experimentos 1, 2, 3 e 4 (Tabela 29). Vale ressaltar que há mais *frames* neutros do que *frames* com expressões, essa informação mostra que devido à rápida execução da EFG durante a sinalização, a rotulação do especialista pode ter sido bastante imprecisa.

Tabela 29 – Detalhes de erros de borda cometidos por classificadores no reconhecimento da EFG **foco**.

Exp.	**F-score**	**Erro total**	**Erro borda**	**Falso neg.**	**Falso neg. na borda**	**Falso Pos.**	**Falso pos. na borda**
1	0,9630	9 (2%)	6 (69%)	4	4 (100%)	5	2 (40%)
1_6	0,9836,	20 (4%)	20 (100%)	9	9 (100%)	11	11 (100%)
2	0,8952	37 (8%)	21 (57%)	22	15 (68%)	15	6 (40%)
2_2	0,9213	31 (7%)	29 (93%)	18	18 (100%)	13	11 (85%)
3	0,8857	121 (9%)	82 (68%)	62	49 (79%)	59	33 (56%)
3_2	0,9022	108 (8%)	65 (60%)	33	26 (79%)	75	39 (52%)
4	0,9120	60 (4%)	22 (37%)	19	16 (84%)	41	6 (15%)
4_2	0 9538	32 (2%)	7 (22%)	0	0 (-)	32	7 (22%)
5	0,8876	728 (53%)	140 (19%)	426	101 (24%)	302	39 (13%)
5_2	0 9321	728 (53%)	140 (19%)	420	106 (25%)	308	34 (11%)

6.10 Expressão Facial Gramatical: Resumo

A série de experimentos e resultados relatados neste trabalho permite delinear algumas considerações. Porém, antes de analisá-las é importante considerar que a modelagem apresentada neste trabalho permitiu a observação de resultados fornecidos por classificadores binários, dedicados à análise e reconhecimento de EFGs de forma individual. A modelagem permitiu a construção de um modelo reconhecedor capaz de identificar uma EFG dentro da execução de uma sentença cujo objetivo semântico era comunicar uma informação que requer a presença de um marcador especial construído por meio da uma expressão facial.

A primeira consideração a ser feita é sobre os resultados de reconhecimento observada para cada EFG. Sob uma análise geral, é factível dizer que o reconhecimento das EFGs foi alcançado com sucesso. As dificuldades encontradas estão concentradas, com maior evidência, na EFG negativa, onde os resultados mais baixos (em termos de F-scores) foram obtidos, e na EFG afirmativa, onde os melhores resultados são os mais fracos quando comparados aos melhores resultados das demais EFGs. Uma das razões para a dificuldade no reconhecimento das EFGs negativas é o fato de sua execução poder assumir diferentes

estilos de configuração de elementos faciais. Outra razão é o fato dessas EFGs assumirem o movimento contínuo da cabeça, inserindo alguma dificuldade na interpretação das relações espaciais entre os pontos que descrevem os elementos da face.

Em relação ao estudo realizado sobre as diferentes formas de representar os elementos das face e suas relações, algumas conclusões podem ser inferidas. Basicamente, apenas as representações com vetores 2 e 4 não apareceram com frequência nos melhores F-scores listados nas tabelas de resultados. Entretanto, houve alguns destaques, ou seja, algumas características que frequentemente apareceram nos melhores resultados. A Tabela 30 resume a análise da representação vetorial, listando as características mais frequentes nos resultados de reconhecimento de cada uma das EFGs.

Tabela 30 – Características que se destacaram nos melhores resultados.

EFG	**Tipo de vetor**	**Informação**	**Ponto de referência**
Interrogativa (qu)	3	Distâncias	Sem / Olho
Interrogativa (sn)	3 / 1	Ângulos	Nariz
Interrogativa (dúvida)	XY / XYZ	Pontos / Ângulos	Sem
Negativa	1	Ângulos	Sem
Afirmativa	1	Distâncias	Sem
Condicional	XY / 1	Distâncias	Nariz
Relativa	1	Distâncias / Ângulos	Sem
Tópico	1 / 3	Distâncias	Sem / Olhos / Nariz
Foco	1	Distâncias / Ângulos	Olhos / Sem

Uma análise alternativa que pode ser realizada é a verificação da relação entre os tipos de vetores de características e o uso ou não de janelas na representação. A Tabela 31 apresenta uma contabilização da frequência de ocorrência dos tipos de vetores e descritores mais presentes nos resultados, em relação ao uso com janelas e sem janelas.

Os resultados obtidos pelos modelos classificadores gerados, considerando os 5 melhores modelos, são bastante similares quando considerados um mesmo experimento e uma mesma EFG. Porém, é possível observar que, de forma geral, o uso de representação com janelas de *frames* leva ao alcance de melhorias no resultado do reconhecimento. Ainda que, em algumas casos, a melhoria tenha sido inferior 4,5%. A Tabela 32 mostra um resumo da melhoria relativa dos melhores resultados que o uso de janelas proporcionou no reconhecimento de cada uma das EFGs. Note que houve uma melhoria em torno de 8% nos

Tabela 31 – Vetores que se destacaram nos melhores resultados.

Melhores Vetores sem Janelas		Melhores Vetores com Janelas	
Vetores	Repetições	Vetores	Repetições
DAO_v3	20	DN_v1 e v3	15
DAQ_v1	18	DAO_v1 e v3	14
DO_v1 e v3	17	DO_v1 e v3	14
AO_v1	15	DAQ	13
DAO_v1	13	XY	12
DN_v3	13	AO_v1	12

experimentos envolvendo as expressões faciais de condicional, tópico e interrogativa (sn), que são expressões que possuem as mesmas modificações faciais (elevação das sobrancelhas e abaixamento da cabeça), com diferença na expressão facial foco que possui modificação nos olhos.

Tabela 32 – Melhoria dos resultados ao utilizar janelas.

EFG	**Menor alteração**	**Maior alteração**	**Janela da maior alteração**
Interrogativa (qu)	0,0014	0,0423	2
Interrogativa (sn)	0,0023	0,0780	6
Interrogativa (dúvida)	0,0146	0,0661	4
Negativa	-0,0103	0,0401	6
Afirmativa	0,0009	0,0386	4
Condicional	0,0075	0,0884	6
Relativa	-0,0064	0,0181	3
Tópico	0,0346	0,0813	6
Foco	0,0165	0,0445	2

Da análise sobre os erros cometidos na borda, é interessante notar que os erros cometidos nas bordas podem ser interpretados como um deslocamento de início ou fim de EFG. E que, ao olho humano, pequenas variações de posição dos elementos da face ocorrida entre dois *frames* pode representar uma situação complexa de análise. Note que se for admitido que o erro de borda pode ser considerado como um fator de precisão de análise do classificador e que, portanto, poderia ser desconsiderado, a acurácia dos classificadores poderia ser considerada mais alta. A Tabela 33 ilustra qual seria a acurácia dos classificadores que alcançaram o maior *F-score* para cada EFG. A Tabela 34 apresenta

a mesma análise considerando os resultados de F-score mais baixos dentre os apresentados neste capítulo, para cada EFG.

Tabela 33 – Melhoria de acurácia representada pela aceitação dos erros de borda, para os experimentos que alcançaram os melhores resultados em termos de **F-score**.

EFG	**Exp.**	**F-score orginal**	**Erro Total original**	**Erro Total final**
Interrogativa (qu)	2_2	0,8988	34 (7%)	**21 (4%)**
Interrogativa (sn)	1_3	0,9412	25 (5%)	**19 (4%)**
Interrogativa (dúvida)	2_11	0,9700	14 (3%)	**5 (1%)**
Negativa	1_3	0,9582	17 (4%)	**10 (2%)**
Afirmativa	1_4	0,8773	33 (9%)	**22 (5%)**
Condicional	1_3	0,9534	17 (3%)	**7 (1%)**
Relativa	2_2	0,9759	9 (1%)	**4 (< 1%)**
Tópico	1_5	0,9544	11 (2%)	**5 (< 1%)**
Foco	1_6	0,9836	20 (4%)	**0**

Tabela 34 – Melhoria de acurácia representada pela aceitação dos erros de borda, para os experimentos que alcançaram os resultados mais baixos dentre os apresentados neste capítulo para cada uma das EFGs.

EFG	**Exp.**	**F-score orginal**	**Erro Total original**	**Erro Total final**
Interrogativa (qu)	5	0,8341	206(13%)	96(6%)
Interrogativa (sn)	3	0,7788	326(19%)	186(11%)
Interrogativa (dúvida)	3	0,8391	290(19%)	208(14%)
Negativa	3_6	0,6760	650(41%)	509(15%)
Afirmativa	3	0,7469	265(25%)	173(12%)
Condicional	3	0,7530	307(15%)	170(11%)
Relativa	3_2	0,8653	161(8%)	106(13%)
Tópico	3	0,8276	155(8%)	35(4%)
Foco	3	0,8857	121(9%)	39(6%)

Por fim, algumas observações adicionais podem ser feitas em relação aos resultados obtidos nos experimentos:

1. Mesmo com informações diferentes o vetor 1 ou o 3 foram os vetores que obtiveram os melhores resultados, lembrando que o vetor 3 possui a informação de profundidade em seus dados.
2. Não foi encontrado evidências de uma melhoria nos resultados ao usar a referência de uma parte da face (olhos ou nariz).
3. Os vetores XY e XYZ não se destacaram em grande parte das expressões faciais.
4. Não houve um tamanho padrão para janelas e alguns experimentos ficaram limitados ao tamanho de janela máximo imposto por frases com presença rápida da expressão facial gramatical.

Apesar da aleatoridade dos valores encontrados nos parâmetros das melhores redes neurais, pode-se constatar que os valores médios para taxa de aprendizado ficaram abaixo de 0,31 e que houve uma tendência em utilizar 13 ou 14 neurônios na camada escondida com uma variação na quantidade de épocas em torno de 50 a 85.

7 Conclusões

O presente trabalho teve por objetivo o desenvolvimento de um conjunto de modelos de reconhecimento de padrões capazes de resolver o problema de reconhecimento de expressões faciais usadas no contexto da Libras, as Expressões Faciais Gramaticais, considerando-as em nível sintático. Os modelos de reconhecimento utilizaram redes neurais Perceptron Multicamadas, treinadas para resolver problemas de classificação binária. Além disso, com o intuito de investigar a influência temporal na representação dos dados, janelas de *frames* foram utilizados como forma de representar o tempo nos experimentos.

Para atingir este objetivo, foram realizados estudos na área da língua de sinais, com foco nas expressões faciais; um estudo sobre técnicas que são aplicadas na área de reconhecimento de expressões faciais e um outro estudo sobre trabalhos correlatos na área de reconhecimento de expressões faciais dentro do escopo de língua de sinais; um estudo sobre técnicas de aprendizado de máquina com Perceptron Multicamadas; coleta de dados utilizando Microsoft Kinect; e o desenvolvimento de algoritmos para extração de características e realização dos experimentos. Em relação aos experimentos e análise dos resultados, pode-se pensar no trabalho em cinco questões a serem investigadas: dependência de usuários no reconhecimento automático das EFs; características que melhor representam as EFs; influência temporal na resolução do problema; influência da informação de profundidade no problema; uso de classificação binária na classificação das EFs.

Uma limitação do trabalho aqui apresentado é que o sinalizador 1, fluente em Libras, é também autor deste trabalho. Embora tenha sido tomado o cuidado de não sofre influências referentes ao conhecimento do problema e conhecimento da resolução do problema de classificação, um viés pode ter acontecido durante a captação dos dados. Porém, também é sabido que, naturalmente, a sinalização do sinalizador 1 é mais bem definida que a sinalização do sinalizador 2, o que é comparável à entonação de voz entre diferentes falas orais. Praticamente todos os experimentos 4 (treino com o segundo sinalizador e teste com o primeiro), tiveram seus resultados superiores aos experimentos 3 (treino com o primeiro sinalizador e teste com o segundo), demonstrando que ao treinar a rede neural com expressões mais sutis, seu desempenho será maior ao testa-las com expressões bem demarcadas. Nos experimentos 5, próximos de uma resolução real, constata-se então que ao misturar diferentes sinalizadores, obtem-se resultados intermediários.

Para identificação dos melhores descritores, foram realizadas combinações com algumas informações (distâncias, ângulos e informação de profundidade), além de alguns testes realizados somente com a informação (x,y) e (x,y,z). Outra possibilidade pensada também foi a necessidade de incluir um ponto referencial em relação à outras medidas. Para isso, foi escolhido utilizar a ponta do nariz e a média da distância entre os olhos. Oito grupos de pontos foram escolhidos por meio do estudo de correlação aplicado aos 100 pontos para seleção menos correlacionados. Os experimentos mostraram que a utilização de um ponto de referência com relação aos outros pontos auxilia na formação de melhores descritores com destaque para os vetores que utilizaram a referência entre os olhos. Com relação a qual informação utilizar, percebe-se que o uso de distâncias e ângulos se destacou na formação dos vetores com presença na maioria dos melhores resultados.

No estudo da dependência temporal, pode-se concluir que a utilização da informação temporal impactou positivamente em praticamente todos os experimentos, com uma janela variando proporcionalmente ao tempo de execução das expressões, como pôde ser observado na expressão facial foco, cuja expressão geralmente é utilizada somente durante a execução de um sinal. Apesar desta constatação, não se identificou uma janela uniforme para todas as expressões, somente uma tendência no tamanho das janelas entre 3 e 6 *frames*. Este impacto positivo ao utilizar uma representação temporal já era esperado devido às informações encontradas nos trabalhos correlatos com foco em língua de sinais.

Por fim, com relação ao uso da informação de profundidade, os vetores que utilizaram a informação de profundidade não se destacaram com relação aos vetores que utilizaram somente a informação (x,y), mostrando que o uso do Microsoft Kinect não se justifica somente pelo fato de proporcionar esta informação, pois existem outras técnicas como *Active Shape Model* (ASM) e *Active Appearance Model* (AAM) que realizam a mesma função de disponibilizar a informação em pixel dos pontos (x,y). No entanto, é necessário um estudo para comparar a precisão destas captações para julgar qual solução seria mais adequada para resolução deste problema.

Pensando nos resultados por expressões faciais, os resultados foram satisfatórios, com algumas dificuldades enfrentadas em algumas expressões, com destaque para expressão facial negativa. Entretanto, grande parte dos experimentos ficaram com F-score acima de 80%, mesmo ao treinar a técnica com um sinalizador e testar com outro, ou treinando e modelando com os dois sinalizadores, demonstrando que a solução, para uma classificação binária, a resolução é factível.

As próximas sessões apresentam um resumo das principais contribuições e sugestões de trabalhos futuros, para construção de uma solução multiclasse e aprimoramento do trabalho atual.

7.0.1 Principais Contribuições

As principais contribuições geradas a partir deste trabalho incluem:

- Revisão da literatura com relação às expressões faciais afetivas, focando nas técnicas utilizadas para extração de características e técnicas utilizadas para classificação automática. Revisão da literatura com relação às expressões faciais dentro do escopo da língua de sinais.
- Projeto, criação e disponibilização[1] do conjunto de dados de EFGs em Libras, proveniente da gesticulação de dois sinalizadores com 5 frases repetidas 5 vezes de 9 expressões faciais, totalizando 90 exemplos de frases de cada expressão facial.
- Estudo da seleção dos pontos menos correlacionados disponibilizados pelo Microsoft Kinect; e apresentação de uma análise detalhada do desempenho de classificadores binários sobre dados faciais obtidos via este sensor.
- Apresentação de um estudo inédito na área de Computação sobre reconhecimento automático de EFGs da Libras[2], com a apresentação de uma análise detalhada sobre a complexidade do reconhecimento automático em cada uma das EFGs.

7.0.2 Trabalhos Futuros

Algumas ideias foram levantadas durante o desenvolvimento desta dissertação de mestrado, visando aprimorar as estratégias utilizadas e construção de uma solução completa.

- **Representação dos dados:**

  Neste trabalho, utilizou pré-processamento (translação dos pontos por meio de um referencial), normalização e correlação para selecionar os grupos de pontos que melhor representassem as informações disponibilizadas pelo dispositivo Microsoft Kinect.

[1] Link para acesso ao acervo: http://archive.ics.uci.edu/ml/datasets/Grammatical+Facial+Expressions
[2] Resultados preliminares foram publicados em (FREITAS et al., 2014)

Uma análise com relação às medidas de distâncias e ângulos mais discriminante pode ser útil na construção de um modelo com maior eficiência e redução de redundância dos vetores de entrada.

- **Diferentes usuários:**

  Percebeu-se que a diferença significativa na marcação expressões faciais durante a sinalização, com duas pessoas com "entonação"bem opostas (sutis e fortes). Acredita-se que uma maior variedade de pessoas para captação dos dados possibilitará uma melhor generalização dos modelos reconhecedores obtidos com uso de técnicas como redes neurais artificiais.

- **Construção de um classificador multiclasse:**

  Esta dissertação de mestrado pode ser utilizada como base para construção de uma solução multiclasse, com uma técnica capaz de identificar expressões faciais gramaticais, afetivas e expressões faciais morfológicas, responsáveis por inferir a intensidade do que se deseja falar.

7.0.3 Considerações Finais

Os estudos na área de língua de sinais são recentes, iniciados em 1960 por Stokoe, com relação a área de reconhecimento automático das expressões faciais neste escopo, não foi identificado nenhum trabalho brasileiro nesta área e poucos trabalhos focados em outras línguas de sinais. Portanto, este trabalho apresenta um estudo inicial sobre o reconhecimento automático das expressões faciais dentro do escopo da Libras.

Por fim, vale ressaltar que este estudo pode ser considerado uma semente para auxiliar o desenvolvimento de uma solução completa de tradução automática da língua de sinais, contribuindo para a quebra na barreira de comunicação entre a comunidade surda e a população ouvinte, utilizando a ciência, para construção de uma ferramenta que contribua para um mundo mais acessível.

Referências[3]

AARONS, D. *Aspects of the syntax of American Sign Language*. Tese (Doutorado) — Citeseer, 1994. Citado na página 24.

AGRIS, U. von; KNORR, M.; KRAISS, K.-F. The significance of facial features for automatic sign language recognition. In: *Automatic Face & Gesture Recognition. FG'08. 8th IEEE International Conference on*. [S.l.: s.n.], 2008. p. 1–6. Citado 6 vezes nas páginas 18, 47, 48, 49, 50 e 51.

AGRIS, U. von; KRAISS, K.-F. Towards a video corpus for signer-independent continuous sign language recognition. *Gesture in Human-Computer Interaction and Simulation, Lisbon, Portugal, May*, 2007. Citado na página 49.

AMARAL, W. M. do. *Sistema de transição da língua brasileira de sinais voltado à produção de conteúdo sinalizado por avatares 3D*. Tese (Doutorado) — Universidade Estadual de Campinas, 2012. Citado na página 26.

ARAN, O. et al. A database of non-manual signs in turkish sign language. In: *Signal Processing and Communications Applications. IEEE 15th*. [S.l.: s.n.], 2007. p. 1–4. Citado na página 49.

ARAN, O. et al. A belief-based sequential fusion approach for fusing manual signs and non-manual signals. *Pattern Recognition*, Elsevier, v. 42, n. 5, p. 812–822, 2009. Citado 2 vezes nas páginas 49 e 51.

ARI, I.; UYAR, A.; AKARUN, L. Facial feature tracking and expression recognition for sign language. In: IEEE. *Computer and Information Sciences. 23rd International Symposium on*. [S.l.], 2008. p. 1–6. Citado 5 vezes nas páginas 47, 48, 49, 50 e 51.

ARROTEIA, J. *O papel da marcação não-manual nas sentenças negativas em Língua de Sinais Brasileira (LSB)*. Tese (Doutorado) — Universidade Estadual de Campinas, 2005. Citado 2 vezes nas páginas 27 e 31.

ARTSTEIN, R.; POESIO, M. Inter-coder agreement for computational linguistics. *Computational Linguistics*, MIT Press, v. 34, n. 4, p. 555–596, 2008. Citado na página 64.

BALOMENOS, T. et al. Emotion analysis in man-machine interaction systems. In: *Machine learning for multimodal interaction*. [S.l.]: Springer, 2005. p. 318–328. Citado na página 42.

BATTISON, R. Phonological deletion in american sign language. *Sign language studies*, v. 5, n. 1974, p. 1–14, 1974. Citado na página 24.

BIOLCHINI, J. et al. *Systematic review in software engineering*. [S.l.], 2005. v. 679, n. 05. Citado na página 20.

BOUCENNA, S.; GAUSSIER, P.; HAFEMEISTER, L. Development of joint attention and social referencing. In: *Development and Learning, IEEE International Conference on*. [S.l.: s.n.], 2011. v. 2, p. 1–6. Citado 2 vezes nas páginas 36 e 37.

[3] De acordo com a Associação Brasileira de Normas Técnicas. NBR 6023.

BRAFFORT, A. A gesture recognition architecture for sign language. In: *Proceedings of the Second Annual ACM Conference on Assistive Technologies*. New York, NY, USA: [s.n.], 1996. (Assets '96), p. 102–109. ISBN 0-89791-776-6. Citado na página 17.

BUENAPOSADA, J. M.; MUÑOZ, E.; BAUMELA, L. Recognising facial expressions in video sequences. *Pattern Analysis and Applications*, Springer, v. 11, n. 1, p. 101–116, 2008. Citado 3 vezes nas páginas 37, 44 e 45.

CAMPR, P.; HRÚZ, M.; TROJANOVÁ, J. Collection and preprocessing of czech sign language corpus for sign language recognition. In: *Language Resources and Evaluation, Proceedings of the Sixth International Conference on*. [S.l.: s.n.], 2008. Citado na página 49.

CARIDAKIS, G.; ASTERIADIS, S.; KARPOUZIS, K. Non-manual cues in automatic sign language recognition. In: ACM. *Pervasive Technologies Related to Assistive Environments, Proceedings of the 4th International Conference on*. [S.l.], 2011. p. 43. Citado 4 vezes nas páginas 18, 47, 48 e 51.

CHANG, C.-Y.; HUANG, Y.-C. Personalized facial expression recognition in indoor environments. In: IEEE. *Neural Networks, International Joint Conference on*. [S.l.], 2010. p. 1–8. Citado 7 vezes nas páginas 16, 37, 43, 45, 46, 68 e 70.

CHEN, J. et al. Facial expression recognition using geometric and appearance features. In: ACM. *Internet Multimedia Computing and Service, Proceedings of the 4th International Conference on*. [S.l.], 2012. p. 29–33. Citado 4 vezes nas páginas 37, 38, 39 e 45.

CHO, M.; PARK, H. Facial image analysis using subspace segregation based on class information. In: SPRINGER. *Neural Information Processing*. [S.l.], 2011. p. 350–357. Citado 3 vezes nas páginas 35, 37 e 44.

DAHMANE, M.; MEUNIER, J. Sift-flow registration for facial expression analysis using gabor wavelets. In: IEEE. *Information Science, Signal Processing and their Applications, 11th International Conference on*. [S.l.], 2012. p. 175–180. Citado 6 vezes nas páginas 37, 38, 41, 45, 46 e 68.

DING, L.; MARTINEZ, A. M. Features versus context: An approach for precise and detailed detection and delineation of faces and facial features. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, IEEE, v. 32, n. 11, p. 2022–2038, 2010. Citado 3 vezes nas páginas 47, 48 e 50.

DUBOIS, J. *Dicionário de lingüística*. [S.l.]: Editora Cultrix, 2001. Citado na página 23.

EKMAN, P. Facial signs: Facts, fantasies, and possibilities. *Sight, sound, and sense*, p. 124–156, 1978. Citado 2 vezes nas páginas 27 e 35.

FERREIRA-BRITO, L. Uma abordagem fonológica dos sinais da lscb. *Espaço: Informativo Técnico-Científico do INES*, p. 20–43, 1990. Citado 3 vezes nas páginas 26, 28 e 30.

FERREIRA-BRITO, L. *Por uma gramática de línguas de sinais*. [S.l.]: Tempo Brasileiro, 1995. Citado na página 24.

FREITAS, F. d. A. et al. Grammatical facial expressions recognition with machine learning. In: *The Twenty-Seventh International Flairs Conference*. [S.l.: s.n.], 2014. Citado na página 98.

FRISHBERG, N. Arbitrariness and iconicity: historical change in american sign language. *Language*, JSTOR, p. 696–719, 1975. Citado na página 23.

GUO, S.; RUAN, Q. Facial expression recognition using local binary covariance matrices. In: IET. *Wireless, Mobile & Multimedia Networks, 4th IET International Conference on*. [S.l.], 2011. p. 237–242. Citado 4 vezes nas páginas 37, 38, 43 e 45.

GWETH, Y. L.; PLAHL, C.; NEY, H. Enhanced continuous sign language recognition using pca and neural network features. In: *Computer Vision and Pattern Recognition Workshops*. [S.l.]: IEEE Computer Society Conference, 2012. p. 55–60. Citado na página 18.

HAN, J.; KAMBER, M. *Data Mining, Southeast Asia Edition: Concepts and Techniques*. [S.l.]: Morgan kaufmann, 2006. Citado na página 21.

HAYKIN, S. *Neural networks and learning machines*. [S.l.]: Pearson Education Upper Saddle River, 2009. Citado na página 56.

HOEY, J.; LITTLE, J. J. Value-directed human behavior analysis from video using partially observable markov decision processes. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, v. 29, n. 7, p. 1118–1132, 2007. Citado 4 vezes nas páginas 36, 37, 42 e 45.

HRÚZ, M.; TROJANOVÁ, J.; ŽELEZNỲ, M. Local binary pattern based features for sign language recognition. *Pattern Recognition and Image Analysis*, Springer, v. 21, n. 3, p. 398–401, 2011. Citado na página 51.

HUANG, X.; LIN, Y. A vision-based hybrid method for facial expression recognition. In: INSTITUTE FOR COMPUTER SCIENCES, SOCIAL-INFORMATICS AND TELECOMMUNICATIONS ENGINEERING. *Ambient media and systems, Proceedings of the 1st international conference on*. [S.l.], 2008. p. 4. Citado 4 vezes nas páginas 35, 37, 40 e 45.

JACK, R. E.; GARROD, O. G.; SCHYNS, P. G. Dynamic facial expressions of emotion transmit an evolving hierarchy of signals over time. *Current Biology*, Elsevier, v. 24, n. 2, p. 187–192, 2014. Citado 3 vezes nas páginas 16, 27 e 35.

JOHO, H. et al. Exploiting facial expressions for affective video summarisation. In: ACM. *Image and Video Retrieval, Proceedings of the ACM International Conference on*. [S.l.], 2009. p. 31:1–31:8. Citado 3 vezes nas páginas 36, 37 e 42.

JR, J. C. W. Signs of change: historical variation in american sign language. *Sign Language Studies*, Gallaudet University Press, v. 10, n. 1, p. 81–94, 1976. Citado na página 23.

KACORRI, H. Models of linguistic facial expressions for american sign language animation. *SIGACCESS Accessibility and Computing*, ACM, n. 105, p. 19–23, 2013. Citado 3 vezes nas páginas 46, 47 e 48.

KANADE, T.; COHN, J. F.; TIAN, Y. Comprehensive database for facial expression analysis. In: *Automatic Face and Gesture Recognition. Proceedings. Fourth IEEE International Conference on*. [S.l.: s.n.], 2000. p. 46–53. Citado na página 39.

KELLY, D. et al. A framework for continuous multimodal sign language recognition. In: ACM. *Multimodal interfaces, International conference on*. [S.l.], 2009. p. 351–358. Citado 6 vezes nas páginas 18, 46, 47, 48, 49 e 51.

KELLY, D. et al. Incorporating facial features into a multi-channel gesture recognition system for the interpretation of irish sign language sequences. In: *Computer Vision Workshops, IEEE 12th International Conference on*. [S.l.: s.n.], 2009. p. 1977–1984. Citado 5 vezes nas páginas 18, 47, 48, 50 e 51.

KELLY, D.; MCDONALD, J.; MARKHAM, C. Recognizing spatiotemporal gestures and movement epenthesis in sign language. In: IEEE. *Machine Vision and Image Processing Conference, 2009. 13th International*. [S.l.], 2009. p. 145–150. Citado na página 51.

KITCHENHAM, B. *Procedures for Performing Systematic Reviews*. [S.l.], 2004. v. 33, n. TR/SE-0401, 28 p. Citado na página 20.

KOSTAKIS, O.; PAPAPETROU, P.; HOLLMÉN, J. Distance measure for querying sequences of temporal intervals. In: ACM. *Pervasive Technologies Related to Assistive Environments, Proceedings of the 4th International Conference on*. [S.l.], 2011. p. 40:1–40:8. Citado 4 vezes nas páginas 18, 47, 48 e 49.

KRNOUL, Z.; HRUZ, M.; CAMPR, P. Correlation analysis of facial features and sign gestures. In: *Signal Processing, IEEE 10th International Conference on*. [S.l.: s.n.], 2010. p. 732–735. Citado 4 vezes nas páginas 47, 48, 49 e 50.

LAJEVARDI, S. M.; HUSSAIN, Z. M. Emotion recognition from color facial images based on multilinear image analysis and log-gabor filters. In: IEEE. *Image and Vision Computing New Zealand, 25th International Conference of*. [S.l.], 2010. p. 1–6. Citado 4 vezes nas páginas 37, 38, 43 e 45.

LEMAIRE, P. et al. Fully automatic 3d facial expression recognition using a region-based approach. In: *Human gesture and behavior understanding, Proceedings joint ACM workshop on*. [S.l.: s.n.], 2011. p. 53–58. Citado 4 vezes nas páginas 37, 38, 40 e 45.

LI, Y.; RUAN, Q.; LI, X. Facial expression recognition based on complex wavelet transform. IET, 2010. Citado 4 vezes nas páginas 37, 38, 43 e 45.

LIDDELL, S. K. *Grammar, gesture, and meaning in American Sign Language*. [S.l.]: Cambridge University Press, 2003. Citado na página 16.

LILLO-MARTIN, D.; QUADROS, R. d. The acquisition of focus constructions in american sign language and língua brasileira de sinais. In: *Boston University Conference on Language Development*. [S.l.: s.n.], 2005. v. 29, p. 365–375. Citado na página 17.

LIU, S.; RUAN, Q.; WANG, Z. An orthogonal tensor rank one discriminative graph embedding method for facial expression recognition. In: *Wireless, Mobile & Multimedia Networks, 4th IET International Conference on*. [S.l.: s.n.], 2011. p. 243–247. Citado 4 vezes nas páginas 37, 38, 43 e 45.

LUCEY, P. et al. The extended cohn-kanade dataset (ck+): A complete dataset for action unit and emotion-specified expression. In: *Computer Vision and Pattern Recognition Workshops, IEEE Computer Society Conference on*. [S.l.: s.n.], 2010. p. 94–101. Citado na página 39.

LYONS, M. et al. Coding facial expressions with gabor wavelets. In: *Automatic Face and Gesture Recognition. Proceedings. Third IEEE International Conference on.* [S.l.: s.n.], 1998. p. 200–205. Citado na página 39.

MADEO, R. C. B. *Máquinas de Vetores Suporte e a Análise de Gestos: incorporando aspectos temporais.* Tese (Doutorado) — Universidade de São Paulo, 2013. Citado na página 22.

MARTINEZ, A. M. The ar face database. *CVC Technical Report*, v. 24, 1998. Citado na página 48.

MCCLELLAND, J. L. et al. Parallel distributed processing. *Explorations in the microstructure of cognition*, v. 2, 1986. Citado na página 56.

MESSER, K. et al. Xm2vtsdb: The extended m2vts database. In: CITESEER. *Audio and video-based biometric person authentication, Second international conference on.* [S.l.], 1999. v. 964, p. 965–966. Citado na página 48.

MICHAEL, N.; METAXAS, D.; NEIDLE, C. Spatial and temporal pyramids for grammatical expression recognition of american sign language. In: *Computers and accessibility, Proceedings of the 11th international ACM SIGACCESS conference on.* [S.l.: s.n.], 2009. p. 75–82. Citado 4 vezes nas páginas 47, 48, 49 e 50.

MPIPERIS, I.; MALASSIOTIS, S.; STRINTZIS, M. G. Bilinear models for 3-d face and facial expression recognition. *Information Forensics and Security, IEEE Transactions on*, v. 3, n. 3, p. 498–511, 2008. Citado na página 40.

NGUYEN, T. D.; RANGANATH, S. Towards recognition of facial expressions in sign language: Tracking facial features under occlusion. In: *Image Processing. 15th IEEE International Conference on.* [S.l.: s.n.], 2008. p. 3228–3231. Citado 2 vezes nas páginas 48 e 51.

NGUYEN, T. D.; RANGANATH, S. Facial expressions in american sign language: Tracking and recognition. *Pattern Recognition*, Elsevier, v. 45, n. 5, p. 1877–1891, 2012. Citado 5 vezes nas páginas 47, 48, 50, 51 e 68.

PATRAS, I.; PANTIC, M. Particle filtering with factorized likelihoods for tracking facial features. In: *Automatic Face and Gesture Recognition. Sixth IEEE International Conference on.* [S.l.: s.n.], 2004. p. 97–102. Citado na página 42.

PETRIDIS, S.; PANTIC, M. Audiovisual discrimination between speech and laughter: why and when visual information might help. *Multimedia, IEEE Transactions on*, v. 13, n. 2, p. 216–234, 2011. Citado 4 vezes nas páginas 36, 37, 42 e 45.

POPA, M.; ROTHKRANTZ, L.; WIGGERS, P. Products appreciation by facial expressions analysis. In: ACM. *Computer Systems and Technologies and Workshop for PhD Students, Proceedings of the 11th International Conference on.* [S.l.], 2010. p. 293–298. Citado 5 vezes nas páginas 36, 37, 38, 42 e 45.

QUADROS, R. M. d.; KARNOPP, L. B. *Língua de sinais brasileira: estudos lingüísticos.* [S.l.: s.n.], 2004. 222 p. Citado 9 vezes nas páginas 16, 17, 23, 27, 28, 29, 30, 32 e 33.

RUDOVIC, O.; PATRAS, I.; PANTIC, M. Coupled gaussian process regression for pose-invariant facial expression recognition. In: *Computer Vision*. [S.l.]: Springer, 2010. p. 350–363. Citado 4 vezes nas páginas 37, 38, 41 e 45.

SAEED, U. Comparative analysis of lip features for person identification. In: ACM. *Frontiers of Information Technology, Proceedings of the 8th International Conference on*. [S.l.], 2010. p. 20:1–20:6. Citado 4 vezes nas páginas 47, 48, 50 e 51.

SARVADEVABHATLA, R. K. et al. Adaptive facial expression recognition using inter-modal top-down context. In: ACM. *Multimodal Interfaces, Proceedings of the 13th international conference on*. [S.l.], 2011. p. 27–34. Citado 3 vezes nas páginas 37, 40 e 45.

SIDDIQUI, M.; LIAO, W.-K.; MEDIONI, G. Vision-based short range interaction between a personal service robot and a user. *Intelligent Service Robotics*, Springer, v. 2, n. 3, p. 113–130, 2009. Citado 3 vezes nas páginas 35, 37 e 45.

SONG, M. et al. A robust multimodal approach for emotion recognition. *Neurocomputing*, Elsevier, v. 71, n. 10, p. 1913–1920, 2008. Citado 3 vezes nas páginas 35, 37 e 45.

STOKOE, W. C. Sign language structure. ERIC, 1978. Citado 3 vezes nas páginas 17, 24 e 25.

SU, M.-C.; ZHAO, Y.-X.; CHEN, H.-F. A fuzzy rule-based approach to recognizing 3-d arm movements. *Neural Systems and Rehabilitation Engineering, IEEE Transactions on*, v. 9, n. 2, p. 191–201, 2001. Citado na página 17.

TAO, H.; HUANG, T. S. Connected vibrations: a modal analysis approach for non-rigid motion tracking. In: *Computer Vision and Pattern Recognition. IEEE Computer Society Conference on*. [S.l.: s.n.], 1998. p. 735–740. Citado na página 42.

TEWS, T.-K. et al. Emotional human-machine interaction: cues from facial expressions. In: *Human Interface and the Management of Information. Interacting with Information*. [S.l.]: Springer, 2011. p. 641–650. Citado 4 vezes nas páginas 35, 37, 38 e 45.

VALENTI, R.; JAIMES, A.; SEBE, N. Facial expression recognition as a creative interface. In: ACM. *Intelligent user interfaces, Proceedings of the 13th international conference on*. [S.l.], 2008. p. 433–434. Citado 3 vezes nas páginas 36, 37 e 42.

VALENTI, R.; JAIMES, A.; SEBE, N. Sonify your face: facial expressions for sound generation. In: ACM. *Multimedia, Proceedings of the international conference on*. [S.l.], 2010. p. 1363–1372. Citado 3 vezes nas páginas 36, 37 e 42.

VIOLA, P.; JONES, M. Rapid object detection using a boosted cascade of simple features. In: *Computer Vision and Pattern Recognition, 2001. Proceedings of the 2001 IEEE Computer Society Conference on*. [S.l.: s.n.], 2001. v. 1, p. I–511. Citado na página 38.

WANG, H. et al. Emotion detection via discriminative kernel method. In: ACM. *Pervasive Technologies Related to Assistive Environments, Proceedings of the 3rd International Conference on*. [S.l.], 2010. p. 7. Citado 6 vezes nas páginas 37, 38, 40, 45, 68 e 70.

WANG, Z. et al. Regularized neighborhood boundary discriminant analysis for facial expression recognition. In: *Wireless, Mobile & Multimedia Networks, 4th IET International Conference on*. [S.l.: s.n.], 2011. p. 248–252. Citado 2 vezes nas páginas 37 e 44.

WERBOS, P. Beyond regression: New tools for prediction and analysis in the behavioral sciences. 1974. Citado na página 56.

WHISSELL, C. The dictionary of affect in language. *Emotion: Theory, research, and experience*, Academic Press, v. 4, n. 113-131, 1989. Citado na página 16.

WU, Q.; SHEN, X.; FU, X. The machine knows what you are hiding: an automatic micro-expression recognition system. In: *Affective Computing and Intelligent Interaction*. [S.l.]: Springer, 2011. p. 152–162. Citado 3 vezes nas páginas 37, 41 e 45.

XIANG, T.; LEUNG, M.; CHO, S. Expression recognition using fuzzy spatio-temporal modeling. *Pattern Recognition*, Elsevier, v. 41, n. 1, p. 204–216, 2008. Citado 3 vezes nas páginas 37, 44 e 45.

XU, M. et al. A vision-based method for recognizing non-manual information in japanese sign language. In: *Advances in Multimodal Interfaces*. [S.l.]: Springer, 2000. p. 572–581. Citado na página 17.

YANG, C.-K.; CHIANG, W.-T. An interactive facial expression generation system. *Multimedia Tools and Applications*, Springer, v. 40, n. 1, p. 41–60, 2008. Citado 2 vezes nas páginas 37 e 45.

YANG, H.-D.; LEE, S.-W. Combination of manual and non-manual features for sign language recognition based on conditional random field and active appearance model. In: IEEE. *Machine Learning and Cybernetics,International Conference on*. [S.l.], 2011. v. 4, p. 1726–1731. Citado 4 vezes nas páginas 47, 48, 50 e 51.

YONG, C. Y.; SUDIRMAN, R.; CHEW, K. M. Facial expression monitoring system using pca-bayes classifier. In: *Future Computer Sciences and Application, International Conference on*. [S.l.: s.n.], 2011. p. 187 –191. Citado 3 vezes nas páginas 37, 42 e 45.

YU, Y.-C.; YOU, S. D.; TSAI, D.-R. Magic mirror table for social-emotion alleviation in the smart home. *Consumer Electronics, IEEE Transactions on*, v. 58, n. 1, p. 126–131, 2012. Citado 5 vezes nas páginas 36, 37, 43, 68 e 70.

ZHANG, L.; TJONDRONEGORO, D.; CHANDRAN, V. Toward a more robust facial expression recognition in occluded images using randomly sampled gabor based templates. In: *Multimedia and Expo, IEEE International Conference on*. [S.l.: s.n.], 2011. p. 1–6. Citado 4 vezes nas páginas 37, 38, 41 e 45.

ZHANG, N.; GENG, X. Facial expression recognition based on local facial regions. In: *Wireless, Mobile & Multimedia Networks, 4th IET International Conference on*. [S.l.: s.n.], 2011. p. 262–265. Citado 4 vezes nas páginas 37, 38, 44 e 45.

ZHAO, S. et al. Video indexing and recommendation based on affective analysis of viewers. In: *Multimedia, Proceedings of the 19th ACM international conference on*. [S.l.: s.n.], 2011. p. 1473–1476. Citado 5 vezes nas páginas 36, 37, 42, 45 e 46.

ZHI, R.; RUAN, Q. Feature extraction using supervised spectral analysis. In: IEEE. *Signal Processing. 9th International Conference on*. [S.l.], 2008. p. 1536–1539. Citado 3 vezes nas páginas 37, 43 e 45.

ZHOU, S.-r.; LIANG, X.-m.; ZHU, C. Support vector clustering of facial expression features. In: IEEE. *Intelligent Computation Technology and Automation, International Conference on*. [S.l.], 2008. v. 1, p. 811–815. Citado 3 vezes nas páginas 37, 41 e 45.

Apêndice A – Revisão Sistemática sobre Análise de Expressões Afetivas: Protocolo e Condução

A.1 Protocolo

•**Objetivos**: Identificar e analisar métodos e técnicas utilizados para a análise de padrões das expressões faciais.

•**Questões de pesquisa**:

1. Quais são os tipos de análises realizadas para as expressões faciais?
2. Quais as técnicas e metódos são aplicados relacionados ao reconhecimento de emoções?
3. Quais são os métodos e técnicas utilizados para extração de características das expressões faciais?
4. Quais os métodos e técnicas utilizados na análise temporal das expressões faciais?

•**Seleção de Fontes**: Os trabalhos devem estar, preferencialmente, disponíveis na internet, em bases de dados científicas. As seguintes bases foram selecionadas para realização das buscas:

1. Biblioteca Digital do IEEE (http://ieeexplore.ieee.org/Xplore/dynhome.jsp)
2. Biblioteca Digital da ACM (http://portal.acm.org/dl.cfm)
3. Biblioteca Digital do Scopus (http://www.scopus.com

•**Idioma dos artigos**: Inglês

•**Palavras-Chave**:

1. Os termos *facial/face expression recognition* com *feature extraction* nos títulos dos artigos.
2. Os termos *human behavior analysis* com *face/facial*
3. Os termos *facial/face expression recognition* com *feature extraction*, excluindo artigos com os termos *facial/face detection, face/facial tracking, face/facial recognition* em seus metadados.

•**Critérios de Inclusão**:

I1 -Artigos que abordem estratégias, métodos e técnicas utilizadas para análise automatizada das expressões faciais.

I2 -Artigos que abordem estratégias, métodos e técnicas utilizados na extração de características das expressões faciais para detectção de emoções.

I3 -Artigos que abordem quais são as características responsáveis por demonstrar as diferentes emoções.

•**Critérios de Exclusão**:

E1 -Artigos que foquem no reconhecimento de faces.

E2 -Artigos que considerem movimentos faciais objetivando a síntese ou reconstrução dos movimentos faciais ou para analisar dados sintéticos.

E3 -Artigos mais antigos do mesmo autor que abordem o mesmo tema, com poucas diferenças entre si.

E4 -Artigos com foco em movimentos da cabeça e de outras partes do corpo, que não seja a face.

E5 -Trabalhos que não tenham sido publicados em conferências ou periódicos (como relatórios técnicos).

- **Estratégia de seleção de dados**: Foram construídas *strings* com palavras-chaves e seus sinônimos, sendo assim submetidas as máquinas de busca. Após a leitura dos resumos, títulos, seções e aplicação de critérios de inclusão e exclusão, o trabalho foi selecionado se confirmado a sua relevância pelo principal revisor. Após definidos os trabalhos definitivamente incluídos, estes foram lidos na íntegra. O revisor fez um resumo esquemático de cada um deles, destacando os métodos para a análise expressões faciais, parâmetros considerados, quando for o caso, o tipo de análise executada e um resumo dos resultados obtidos para uma comparação futura.
- **Síntese dos dados extraídos**: Após a leitura e o resumo esquemático, foi elaborado um relatório com uma análise quantitativa dos trabalhos, norteado pelos passos enumerados abaixo.

1.Identificação das vantagens e desvantagens de cada método.

2.Qual técnica utilizada?

3.Quantas características foram analisadas da face?

4.Quantas emoções foram consideradas?

5.Ambiente artificial ou natural?

6.Seleção dos marcadores da face de forma automatizada ou manual?

7.Qual software ou *framework* utilizado?

A.2 Condução

Após planejar a revisão sistemática e definir um protocolo para ela, a condução dos passos definidos consistiu em, resumidamente: submeter as *strings* às máquinas de busca da ACM, IEEE e Scopus; submeter os artigos retornados destas buscas aos critérios de inclusão e exclusão; e análisar de forma completa cada artigo.

As buscas foram realizadas entre os dias 20 e 31 de outubro de 2012. Nestas buscas, 72 artigos foram encontrados, sendo que após a aplicação dos critérios de inclusão e exclusão, 35 artigos foram selecionados para leitura e análise completa, sendo 16 artigos do IEEE, 11 artigos da ACM e 8 do Scopus.

Apêndice B – Revisão Sistemática sobre Análise de Expressões Gramaticais: Protocolo e Condução

B.1 Protocolo

•**Objetivos**: Identificar e analisar métodos e técnicas utilizadas para a análise das expressões faciais das Línguas de Sinais.

•**Questões de pesquisa**:

1. Quais são os tipos de análises realizadas para as expressões faicias das línguas de sinais?
2. Quais são as técnicas e métodos utilizados para extração das características das expressões faciais?
3. Quais são as EFs usualmente trabalhadas no contexto das línguas de sinais?
4. Quais são as técnicas e métodos utilizadas considerando o aspecto temporal durante o resconhecimento das expressões faciais tendo em vista o discurso nas língua de sinais?

•**Seleção de fontes**:

•**Seleção de Fontes**: Os trabalhos devem estar, preferencialmente, disponíveis na internet, em bases de dados científicas. As seguintes bases foram selecionadas para realização das buscas:

1. Biblioteca Digital do IEEE (http://ieeexplore.ieee.org/Xplore/dynhome.jsp)
2. Biblioteca Digital da ACM (http://portal.acm.org/dl.cfm)
3. Biblioteca Digital do Scopus (http://www.scopus.com)

•**Idioma dos artigos**: Inglês

•**Palavras-Chave**:

1. Os termos *sign language* E *facial/face expression*.
2. O termo *emotion* E *sign language*.
3. O termo *sign language* E *non-manual/nonmanual*

•**Critérios de Inclusão:**

I1 -Artigos que abordem estratégias, métodos e técnicas utilizadas para análise automatizada das EFs das LS.

I2 -Artigos que abordem estratégias, métodos e técnicas utilizadas extração das características das EF.

•**Critérios de Exclusão:**

E1 -Artigos que foquem em reconhecimento de face.

E2 -Artigos mais antigos do mesmo autor que abordem o mesmo assunto com poucas diferenças entre si.

E3 -Artigos que foquem em reconhecimento de sinais.

E4 -Trabalhos que não tenham sido publicados em conferências ou periódicos, como relatórios técnicos.

E5 -Trabalhos que não se encaixem nos critérios de inclusão.

- **Estratégia de seleção de dados**: Foram construídas *strings* com palavras-chaves e seus sinônimos, sendo assim submetidas as máquinas de busca. Após a leitura dos resumos, títulos, seções e aplicação de critérios de inclusão e exclusão, o trabalho foi selecionado se confirmado a sua relevância pelo principal revisor. Após definidos os trabalhos definitivamente incluídos, estes foram lidos na íntegra. O revisor fez um resumo esquemático de cada um deles, destacando os métodos para a análise expressões faciais, parâmetros considerados, quando for o caso, o tipo de análise executada e um resumo dos resultados obtidos para uma comparação futura.
- **Síntese dos dados extraídos**: Após a leitura e o resumo esquemático, foi elaborado um relatório com uma análise quantitativa dos trabalhos, norteado pelos passos enumerados abaixo.

1.Identificação das vantagens e desvantagens de cada método.
2.Qual técnica utilizada?
3.Quantas características foram analisadas da face?
4.Quantas EFs foram consideradas?
5.Tipo de dados? Primários ou secundários?
6.Seleção de pontos responsáveis pelo reconhecimento de forma automatizada ou manual?
7.Qual o software ou *framework* utilizado?

B.2 Condução

As *strings* foram submetidas às maquinas de busca da IEEE, ACM e Scopus. Esta busca foi realizada entre os dias 20 e 27 de maio de 2013 e foram encontrados 82 artigos. Após a aplicação das regras de inclusão e exclusão, restaram apenas 14 artigos, sendo 7 do IEEE, 6 da ACM e 1 do Scopus.

Apêndice C – Pontos fornecidos pela Aplicação de Aquisição de Dados

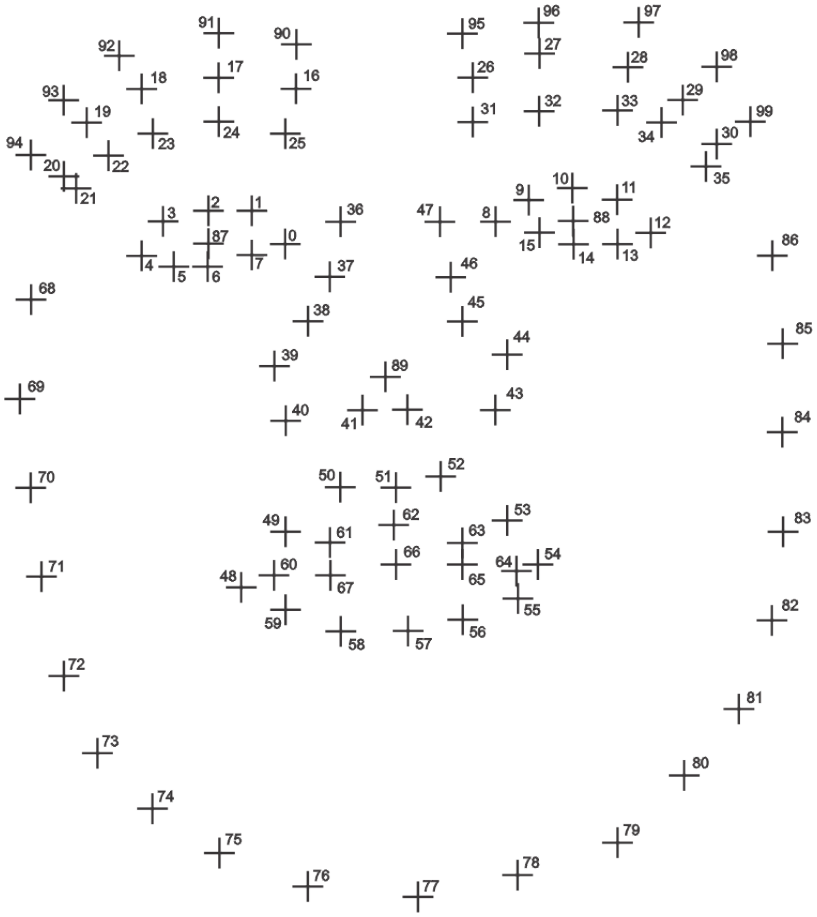


Figura 20 – Visão detalhada dos 100 pontos fornecidos pela aplicação de aquisição de dados.