



Universidade Católica Dom Bosco
Curso de Bacharelado em Engenharia de Computação

**Reconhecimento de Gestos Utilizando
Modelos de Markov Ocultos**

Jéssica Barbosa Dias
Kleber Padovani de Souza

Orientador: Prof. Dr. Hemerson Pistori

Relatório final apresentado à Coordenação do Curso de Bacharelado em Engenharia de Computação da Universidade Católica Dom Bosco como parte dos requisitos para a obtenção do título de Bacharel em Engenharia de Computação.

UCDB - Campo Grande - MS - Novembro/2006

Dedicatórias e Agradecimentos dos autores

Cada ser-humano envolvido no desenvolvimento de um trabalho possui forças, inspirações, medos e dificuldades específicas e particulares. Como este trabalho foi desenvolvido em dupla, decidimos dividir este espaço, a fim de que cada um pudesse expressar suas sinceras dedicatórias e seus agradecimentos a todos que contribuíram de alguma maneira com este projeto durante toda a árdua caminhada.

Dedicatória

A Deus, aos meus irmãos Ronald e Ingrid e principalmente aos meus pais Jurandir e Deolinda, que me proporcionaram esta oportunidade.

Agradecimentos

A minha família pelo incentivo, apoio, compreensão e confiança em mim depositada e que mesmo em dificuldades nunca solicitaram minha desistência.

Ao orientador Hemerson Pistori, pela paciência, pelos ensinamentos e a competência ao conduzir este projeto.

Aos amigos que sempre deram força para prosseguir, em particular a Gilmara que quando necessitei sempre se fez presente.

Aos amigos feito na universidade como Willian, Laura, Anderson Pacheco e os demais que ajudaram, cada um de sua maneira e também que me fizeram rir ao longo desses cinco anos, não podendo esquecer de agradecer ao Kleber uma pessoa especial que me deu a oportunidade de ter participado deste projeto, me ensinando e auxiliando e a Lia Nara que contribuiu na etapa final do projeto mesmo não tendo interesse, apenas pela amizade. E a todos que contribuíram para minha formação pessoal, profissional e para que este projeto fosse concluído com sucesso.

Jéssica Barbosa Dias.

Antes de agradecer a qualquer pessoa, gostaria de agradecer a Deus, que, mesmo diante de toda minha ingratidão e incontáveis falhas, está presente e me dá muita força para continuar a caminhada. Nesses cinco anos, como em vários outros, também enfrentei muitos momentos conturbados em minha vida, mas Ele sempre foi o amigo fiel que pude contar para me ajudar a levantar nos momentos em que eu caí, até mesmo quando não pedi Sua ajuda. A você, meu grande amigo e testemunha do meu sofrimento diário, meu agradecimento sincero e especial.

Agradeço aos meus pais pelo apoio, confiança, amor, exemplo e todas as coisas positivas que me passaram e que me fizeram hoje o homem que sou. Muito obrigado pelos valores, pelos conselhos e principalmente pela vida. Sei que vocês vibram com esta conquista tanto como eu, talvez até mais, e é a vocês que dedico essa vitória e todo o esforço desses cinco anos. Sou muito grato a vocês! Agradeço também ao Kleyber, meu irmão, pelo apoio que pôde me dar durante esse período.

Agradeço às minhas primas, Lidiane, Lílian e Liliane, que são parte de minha vida e estão entre as pessoas que mais amo e gosto de estar perto. Obrigado pela companhia durante os vários fins de semana, pelas risadas, pelos lanches de sábado à noite, pelos almoços de domingo... Enfim, são muitas coisas a agradecer. Sou agradecido também ao tio "Verso" e tia "Raxel", pelo apoio e carinho que me dedicaram sempre que precisei. Sou muito grato à Dra. Jadeusi, que foi um dos melhores presentes que Deus me deu. Ela sabe o quanto ela foi, e é, especial na minha vida. Obrigado pelo apoio de sempre!

Obrigado ao professor Hemerson pela confiança, que por vezes não foi retribuída, pelos conselhos, pela amizade e pelo apoio que sempre nos ofereceu. Agradeço mais pelo seu esforço para tornar o grupo de pesquisa cada vez mais forte e reconhecido no meio acadêmico e pela valorização e respeito que ele tem com todos os integrantes do grupo, independente do cargo exercido dentro da instituição.

Agradeço ao grande professor e amigo, Amaury Antônio Castro Júnior, que, durante a maior parte desses cinco anos, acompanhou essa jornada, sempre agüentando "um certo mala" em suas aulas e oferecendo um ombro amigo quando os "chorões" pediam arrego. Amaury, você sabe que te admiro muito como pessoa e profissional. É uma pena que alguns não tenham lhe dado o devido valor, mas onde estiver você se destacará por próprio mérito. Obrigado... e parabéns pela sua competência!

Agradeço ao Sr. Willian Paraguassu Amorim, vulgo "Putinha", por sempre me ajudar a "vacalhar" o e-mail e o MSN do pessoal e sempre tirar o corpo fora e me deixar com a culpa; obrigado por nunca me avisar que já havia ido pra faculdade quando pedia pra eu te dar carona; obrigado por sempre inventar uma desculpa mais criativa que a outra para não ir à nenhuma confraternização do pessoal da sala; obrigado por reclamar que o ventilador do laboratório seca seu olho nos dias mais quentes do ano; obrigado por dizer que eu conto sempre as mesmas piadas sem graça; e... obrigado pela alegria que nos transmitiu e por todos os momentos hilários que nos proporcionou durante todo o curso, devido à

sua maneira engraçada e "relax" de ver a vida e enfrentar os problemas do dia-a-dia.

Agradeço à nossa copiadora oficial Jessy Black, que abandonou o cargo recentemente, por ter responsabilidade por nós todos e copiar a matéria que ninguém copiava. Obrigado, Jessy, pelo companheirismo, pela dedicação, pelo empenho, pelas risadas, pela sinceridade, pela amizade e, é claro, por ter me emprestado vários reais nos primeiros anos. Jessy, obrigado por ter sido parte desse projeto e por ter se empenhado durante todo esse ano.

Obrigado à Laura "Popozão" por esmagar meu freegel's, com seu jeitinho "carinhoso" e "delicado" de ser; obrigado pelo convívio durante todo esse tempo; obrigado pelas brigas, pelos risos, pelas chacotas, pelos pagodes... por tudo.

Obrigado aos trigêmeos, Bruno, Jonathan e Wesley, pelo apoio durante esse período de convívio na pesquisa, pela bagunça, pelo tombo que o Bruno levou na SBC, pelos doc's do SVN e, é claro, pela troca de conhecimentos.

Um "obrigação" pra Lia Nara Balta Quinta, a Piazinha, que na reta final vestiu a camisa e se empenhou em ajudar a gente como se fosse parte do projeto. Lia, muito obrigado MESMO pela sua ajuda, pela sua companhia e pelas discussões... E obrigado também à sua mãe, que nos emprestou o Santo das Causas Impossíveis que deu "aquele" empurrãozinho nos 45 do segundo tempo. Agradeço às colaboradoras Elaine e Dolores, que contribuíram de maneira ímpar, cedendo suas imagens ao projeto.

Agradeço às nossas meninas, Grazy Arruda Pancadão e CCET Neêmia, por nos fazerem companhia e contribuírem para que os "nerds", termo atribuído carinhosamente pela Grazy, não se estressassem muito com a rotina de pesquisa.

Agradeço ao professor Mauro, pelas traduções de textos para o Inglês, pelos auxílios como professor, e principalmente por não tomar o chá que é servido aos professores e mesmo assim não compartilhá-lo conosco. Muito Mate Leão, pra você!

Agradeço à minha turma, que por cinco anos teve a extensa paciência de me tolerar. Espero que vocês também, mas eu me diverti muito. Agradeço a todos os professores que tiveram o prazer, ou o desprazer, de me ter como aluno. Gostaria de dizer que todo o conhecimento que pude absorver de suas aulas têm sido de grande valia para mim, e espero crescer ainda mais profissionalmente utilizando essa bagagem que vocês me deram. Agradeço também à toda a "galera do fervo" do GPEC, que sempre está se divertindo e compartilhado os conhecimentos adquiridos, proporcionando o crescimento geral do grupo.

Um super agradecimento ao meu parceiro fiel, que nunca "negou fogo" nesse laboratório e tornou possível a realização desse trabalho: meu PC!

Enfim, a todos, inclusos ou não nesses agradecimentos, minha única intenção é dizer a vocês o meu sincero...



Valeu!!!

Resumo

Os modelos de Markov ocultos (HMM) têm sido utilizados com frequência e sucesso em sistemas reconhecedores de fala, sendo isto uma das inspirações para sua aplicação nas pesquisas de reconhecimento de gestos. O reconhecimento de gestos via computação traz benefícios importantes à comunidade integrante do mundo digital, em especial, aos portadores de dificuldades auditivas e motoras, que possuem certos impedimentos no uso dos dispositivos tradicionais de interação homem-máquina. Neste trabalho será apresentado um sistema baseado em modelos de Markov ocultos capaz de reconhecer gestos humanos em seqüências de imagens, integrado a uma plataforma de apoio ao desenvolvimento de aplicações guiadas por sinais visuais.

Abstract

Hidden Markov Models (HMM) have been frequently used in speech recognition, what suggested their possible use in gesture recognition. Traditional human-machine interfaces are not suitable for people with certain disabilities, specially auditory and motor disabilities, and the use of computers to automatically recognize gestures can bring important benefits to this community. This work presents a HMM-based system capable of recognizing human gestures in sequences of images, integrated in a platform to support the development of visually guided applications.

Conteúdo

1	Introdução	10
2	Fundamentação Teórica	14
2.1	Visão Computacional e Reconhecimento de Padrões	14
2.2	Probabilidade e estatística	15
2.3	Língua de Sinais	18
3	Modelo de Markov Oculto	21
3.1	Elementos do HMM	22
3.2	Problemas Básicos	24
3.2.1	Problema de avaliação	24
3.2.2	Problema da busca da melhor seqüência de estados . .	28
3.2.3	Problema de treinamento	28
4	Desenvolvimento	31
4.1	Modelagem, Captura e Análise dos Gestos	31
4.2	Implementações	41
4.2.1	Análise Visual e Obtenção de Componentes Iniciais dos HMMs	42
4.2.2	Segmentação e Extração de Atributos	43
4.2.3	Discretização de valores e Geração de Observações . . .	44
4.2.4	Obtenção das componentes finais dos HMMs	45
4.2.5	Classificação das observações	45
4.2.6	Diagramas de funcionamento	46
5	Experimentos	49
6	Considerações Finais	54
A	Tabelas de Atributos das Características	57
	Referências Bibliográficas	60

Capítulo 1

Introdução

Cada vez mais os computadores têm se tornando parte da vida das pessoas, tanto nas atividades pessoais quanto nas profissionais. A facilidade na utilização destes equipamentos é uma característica essencial em seus projetos, pela simples razão que dispositivos de difícil interação são menos utilizados e, geralmente, causam um desperdício de tempo ao usuário devido à necessidade de treinamento [11].

De forma geral, a interação entre pessoas e computadores (interface homem-máquina) atualmente ocorre através dos dispositivos de entrada de dados habituais, como o mouse e o teclado. Através da utilização de dispositivos óptico-eletrônicos, como as *webcams*, é possível enriquecer este paradigma de comunicação. A análise e reconhecimento dos gestos, posturas e expressões humanas podem auxiliar muito a interface homem-máquina.

Segundo o Censo Demográfico 2000 do Instituto Brasileiro de Geografia e Estatística, cerca de 937 mil brasileiros são tetraplégicos, paraplégicos ou hemiplégicos permanentes, e mais de 5 milhões são incapazes, ou possuem alguma ou grande dificuldade permanente de ouvir. Considerando o fato de existir uma parcela significativa de pessoas atualmente integral, ou parcialmente impedidas de utilizar os computadores da forma convencional, um outro benefício muito importante do reconhecimento de gestos pelos computadores é a sua contribuição na inclusão digital desta categoria de pessoas que possuem necessidades especiais. Um sistema capaz de identificar comandos do usuário por meio da movimentação do globo ocular poderia ser utilizado por tetraplégicos como um meio de comunicação entre eles e os dispositivos, o que os incluiriam no mundo digital considerando suas limitações.

Por fim, existe uma grande quantidade de aplicações que podem ser desenvolvidas utilizando comunicação homem-máquina guiadas por sinais visuais. Porém, um dos empecilhos para a propagação do desenvolvimento deste tipo de aplicativo é a complexidade de implementação da interface. Para

suprir esta necessidade foi criado um projeto, denominado SIGUS [15], que tem como objetivo o desenvolvimento de um ambiente computacional, de códigos-fonte abertos, que auxilie o desenvolvimento de aplicações que sejam guiadas por interfaces não-convencionais, mais especificamente, guiadas por sinais visuais, como movimentos da mão e da face.

Atualmente, esta ferramenta possui várias implementações de técnicas e algoritmos da área de visão computacional e reconhecimento de padrões, dentre elas algoritmos de segmentação, extração de características e classificação de imagens. Através dela, foram implementados alguns programas para análise da eficácia da aplicação destas teorias, e foram obtidos resultados satisfatórios. Um exemplo de implementação é um protótipo de editor de textos para um pequeno sub-conjunto da Língua Brasileira de Sinais (LIBRAS), em que o usuário gestualiza o símbolo referente a uma letra do alfabeto LIBRAS em frente à câmera, que captura as imagens; em seguida, o computador processa as imagens e traduz a postura para a linguagem tradicional, exibindo, por fim, a letra informada pelo usuário.

Contudo, na LIBRAS, assim como em outras línguas corporais, a comunicação não se limita apenas em posturas, que são imagens sem movimento, mas também em gestos completos, o que torna o processo de reconhecimento computacional mais complexo. Na classificação científica de gestos comunicativos das mãos e dos braços, a língua de sinais é geralmente considerada como a mais estruturada, dentre as várias categorias de gestos [13]. Entretanto, existem estudos na área de visão computacional que podem auxiliar este processo e, conseqüentemente, obter os resultados desejados. Para este problema específico, uma das alternativas que auxiliaria sua solução seria a utilização de modelo de Markov oculto (*Hidden Markov Model - HMM*).

Embora inicialmente introduzido e estudado no fim dos anos 60, os modelos de Markov ocultos têm se popularizado cada vez mais nos últimos anos. Uma das razões é porque os modelos, quando utilizados corretamente, funcionam muito bem na prática para muitas aplicações importantes [17]. Frequentemente utilizadas para o reconhecimento de fala, as técnicas de HMM foram inseridas há pouco tempo na área de visão computacional, sendo os trabalhos mais recentes limitados ao reconhecimento de manuscritos, como descrito em [6]. Porém, alguns trabalhos utilizaram-na para o reconhecimento de gestos e obtiveram resultados satisfatórios.

Como exemplo, em [21] é desenvolvido um sistema para o reconhecimento da Língua Americana de Sinais (*American Sign Language - ASL*). Neste trabalho, utiliza-se um dicionário de 40 palavras, constituído de pronomes pessoais, verbos, substantivos e adjetivos, para criar 494 sentenças de cinco palavras cada, que são gestualizadas pelo usuário do computador, com auxílio de luvas coloridas. Utilizando uma regra gramatical esse sistema conseguiu

uma taxa de acerto de 97%; e sem a gramática, foi alcançada uma taxa de 91% de acerto.

Donald Tanguay [22] também executou um trabalho que consiste no desenvolvimento, implementação e experimentação de um sistema para criação de gestos, treinamento de HMMs e reconhecimento dos gestos através dos HMMs. Em seus experimentos, foi realizado o reconhecimento de movimentos do mouse, em que foi obtida uma taxa de acerto de 97.5%, testando-o com 40 exemplos; e 88.75% com 80 exemplos.

Em [1] é descrito o reconhecimento de gestos da cabeça humana através do uso dos HMMs. Em um dos experimentos deste trabalho, foram utilizadas para o treinamento 28 seqüências, obtidas de 5 pessoas; e como conjunto de testes foram 33 seqüências, obtidas de 6 outras pessoas. Foram utilizados 4 tipos diferentes de expressões: (a) sim; (b) não; (c) talvez; e (d) olá. A Tabela 1.1 exhibe a matriz de confusão dos resultados obtidos do conjunto de testes, com base no conjunto de treinamento acima descritos. As linhas da tabela representam o resultado esperado para cada entrada a ser classificada pelo sistema, e as colunas o resultado obtido como resposta do sistema. Ou seja, a intersecção entre valores iguais de linha e coluna corresponde às respostas corretas do sistema, e as outras, conseqüentemente, às erradas. Como exemplo, dos 18 gestos que deveriam ser classificados como SIM, 13 foram classificados corretamente, 2 foram classificados como TALVEZ e 3 como OLÁ.

	SIM	NÃO	TALVEZ	OLÁ
SIM	13		2	3
NÃO		5		
TALVEZ	1	1	4	
OLÁ	2			2
TOTAL	16	6	6	5

Tabela 1.1: Matriz de confusão dos resultados obtidos.

O objetivo deste trabalho foi estudar e desenvolver aplicações sobre o modelo de Markov oculto, baseadas na plataforma SIGUS, aplicadas ao reconhecimento de gestos humanos através da visão computacional, utilizando suas teorias e analisando sua eficácia através da aplicação em bancos de imagens. Com isso, pretende-se obter como resultado um sistema reconhecedor de gestos eficaz através de visão computacional, que possa ser adaptado à plataforma SIGUS, podendo beneficiar diretamente o grupo de pessoas que possuem necessidades especiais, facilitar a interação homem-máquina e gerar

um material que possa contribuir para pesquisas futuras.

Primeiramente, foram estudados trabalhos relacionados com a área, como [5, 8, 12, 21, 24], para obtenção de conhecimentos e base teórica para a execução do projeto. Em seguida, foram selecionados de [18, 19] os gestos a serem classificados pelo sistema. Para realizar o treinamento do sistema foram realizadas as gravações das imagens dos gestos selecionados, que posteriormente foram processadas e utilizadas como fonte de informação para alimentar o sistema. Para cada gesto selecionado foi criado um modelo de Markov oculto. Após as etapas anteriores, foi implementada a aplicação reconhedora de gestos, com base na plataforma SIGUS. Em seguida, começaram a ser realizados os experimentos no sistema, testando sua eficácia em diferentes situações.

A descrição deste trabalho foi organizada em 5 etapas. No capítulo 2, são fornecidas informações teóricas de base, para que o leitor possa compreender de maneira mais simples os capítulos posteriores do trabalho. O capítulo 3 descreve o conceito da técnica central do trabalho, os modelos de Markov ocultos (HMM). O capítulo seguinte explica detalhadamente a forma em que o conceito foi aplicado ao reconhecimento de gestos, os passos para a execução da aplicação e descreve as implementações criadas. Por fim, o capítulo 5 apresenta os experimentos e resultados, faz uma análise breve e geral sobre o tema discutido neste trabalho e sugere possíveis trabalhos futuros.

Capítulo 2

Fundamentação Teórica

2.1 Visão Computacional e Reconhecimento de Padrões

A visão computacional, ou visão de máquina, é uma área de conhecimento que se dedica a desenvolver teorias e métodos voltados à extração automática de informações “úteis” contidas em imagens, sendo que a utilidade de uma informação é altamente dependente da aplicação. Por exemplo, para uma indústria de manufatura, que fabrica determinado produto, uma informação útil pode estar relacionada com alguma diferença de cor, ou forma padrão, esperada para um produto sem defeitos.

Basicamente, sistemas de visão computacional capturam imagens através de dispositivos óptico-eletrônicos, como câmeras e filmadoras digitais, e buscam produzir descrições úteis das informações contidas nas imagens. Essas descrições podem ser utilizadas, por exemplo, na classificação de objetos ou no controle automático de algum dispositivo atuador, como um braço robótico ou uma rede de esteiras rolantes em uma linha de montagem. Estas tarefas relativamente simples de serem realizadas por seres humanos, como diferenciar chaves de fenda de chaves alemãs em uma linha de montagem, apresentam-se como grandes desafios para sistemas automáticos de visão computacional.

Embora o reconhecimento de padrões possa ser aplicado a problemas sem qualquer relação com imagens e visão, existe uma rica intersecção entre essa área e a área da visão computacional, uma vez que o reconhecimento de padrões é uma importante etapa em boa parte dos problemas de visão computacional. Reconhecimento de padrões é a disciplina científica cujo objetivo é a criação de teorias e técnicas que permitam a classificação de objetos, ou padrões, dentre um conjunto de categorias ou classes [23]. Dependendo da

aplicação, esses objetos podem ser imagens, seqüências de caracteres, sons ou qualquer outro tipo de sinal, geralmente digitalizado, capturado através dos sensores de um sistema computacional. Um exemplo que se encaixaria nessa intersecção entre as duas áreas, seria um sistema de tradução de linguagem corporal para natural, em que o usuário informaria comandos ao computador através de suas imagens capturadas em uma webcam. Existem trabalhos relacionados que têm como alvo a construção de sistemas de reconhecimento de língua de sinais para deficientes auditivos [14].

Nestes sistemas geralmente existem duas etapas importantes: o reconhecimento de posturas e o reconhecimento de gestos. Este reconhecimento se baseia na busca pelo modelo que possui características que melhor se assemelham com os parâmetros extraídos da imagem corrente. Posturas são sinais que não envolvem movimentação, ou seja, são estáticos, e, por este motivo, tornam sua modelagem mais simples comparada aos gestos. A modelagem dos gestos inclui informações temporais e análise de seqüências de imagens. A grande maioria dos trabalhos de reconhecimento de gestos utiliza técnicas adaptadas de reconhecimento da fala, como as baseadas em cadeias de Markov [14].

2.2 Probabilidade e estatística

A probabilidade está presente em várias áreas de estudo, tais como economia, biologia, engenharia, física e até mesmo em nosso cotidiano. Sempre que nos deparamos com situações que envolvem “chances” temos a probabilidade implícita nelas. Como exemplo, em jogos de sorte existem as probabilidades do jogador ganhar e do jogador perder; em gestações existem as probabilidades do feto ser do sexo masculino e do feto ser do sexo feminino; nas bolsas de valores existem as probabilidades dos valores das ações aumentar, como também diminuir, devido a determinadas situações.

Um dos conceitos de *probabilidade* é definido pelo quociente entre o número de casos favoráveis e o número total de casos possíveis em uma experiência. Por exemplo, considere o exemplo de uma casa com 10 pessoas, sendo 3 mulheres, 5 crianças e 2 homens. A probabilidade de um indivíduo tocar a campainha e ser atendido por uma criança é de 50%, $\frac{5}{10} = 0.5$, pois são 5 casos favoráveis dentre 10 casos possíveis de serem observados. Da mesma forma que temos 30% de chance do indivíduo ser atendido por uma mulher, dado que existem 3 mulheres, que são 3 observações favoráveis dentro de uma sala com 10 possíveis observações. Portanto, a probabilidade de um evento ocorrer é descrito pela Equação 2.1, em que A é o evento desejado, que, no caso acima, é a probabilidade de escolher uma criança, uma mulher

ou um homem.

$$P(A) = \frac{\text{Número de casos favoráveis}}{\text{Número de casos possíveis}} = \frac{N_A}{N} \quad (2.1)$$

Diz-se que dois eventos são *independentes* quando a ocorrência de um não interfere na ocorrência do outro [9]. Utilizando o mesmo exemplo, considere que 4 pessoas da sala comeram carne bovina no almoço, 3 comeram peixe e as outras 3 frango. A probabilidade do indivíduo ser atendido por uma criança e a probabilidade dele ser atendido por uma pessoa que almoçou frango podem ser considerados como eventos independentes. O cálculo da probabilidade da ocorrência de dois eventos independentes dá-se pela seguinte fórmula:

$$P(A \text{ e } B) = P(A).P(B) \quad (2.2)$$

Ou seja, a probabilidade do indivíduo ser atendido por uma criança que tenha almoçado frango é:

$$P(A \text{ e } B) = P(A).P(B) = \frac{5}{10} \cdot \frac{3}{10} = 0.15 = 15\% \quad (2.3)$$

em que A é a probabilidade de ser atendido por uma criança e B de ser atendido por uma pessoa que tenha almoçado frango.

Dois eventos são *mutuamente exclusivos* quando a ocorrência de um elimina a possibilidade de ocorrência do outro [7]. Por exemplo, dado que a probabilidade do indivíduo ser atendido por um homem e a probabilidade do indivíduo ser atendido por uma mulher sejam os eventos A e B, respectivamente. É impossível a ocorrência dos dois eventos simultaneamente, pois ele pode ser atendido por um homem ou por uma mulher e nunca pelos dois, considerando que apenas uma pessoa atende ao indivíduo por vez. O valor da probabilidade de ocorrência de eventos mutuamente exclusivos é dado pela seguinte fórmula:

$$P(A \text{ ou } B) = P(A) + P(B) \quad (2.4)$$

Nesse sentido, a probabilidade do indivíduo ser atendido por um homem ou por uma mulher é:

$$P(A \text{ ou } B) = P(A) + P(B) = \frac{2}{10} + \frac{3}{10} = 0.05 = 5\% \quad (2.5)$$

Existem necessidades de cálculos de probabilidade mais complexos, como é o caso da *probabilidade condicional*, que é a probabilidade de um evento ocorrer dada a ocorrência de outro evento. Continuando no exemplo citado, vamos supor ainda que, dentre os dez, um homem, duas mulheres e uma

criança são loiros; um homem e duas crianças são morenos; e uma mulher e duas crianças são ruivas. A probabilidade do indivíduo ter sido atendido por uma criança, dado que a pessoa que atendeu é loira, é a seguinte:

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \quad (2.6)$$

em que A é a probabilidade da pessoa ser uma criança e B a informação que a pessoa é loira; $P(A|B)$ é a probabilidade desejada, e lê-se como a probabilidade de A , dado B ; e $P(A \cap B)$ é o número de casos favoráveis em que A e B ocorrem juntamente. No exemplo, $P(A \cap B)$ é a probabilidade de encontrar uma criança loira entre todos da casa.

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{\frac{1}{10}}{\frac{4}{10}} = 0.25 = 25\% \quad (2.7)$$

Essa fórmula é aplicada para o cálculo de *probabilidade condicional para eventos dependentes*, ou seja, a ocorrência do evento dado interfere na probabilidade de ocorrência do outro evento. Quando isto não ocorre estamos lidando com *probabilidade condicional para eventos independentes* e, como o evento desejado não depende do evento anterior, a probabilidade se resume a ele próprio, conforme é exibido na Equação 2.8.

$$P(A|B) = P(A) \quad (2.8)$$

Por fim, uma última probabilidade pode ser citada, que é a probabilidade da pessoa que atender o indivíduo ser uma mulher ou ser uma pessoa loira. Ao contrário do exemplo citado para eventos mutuamente exclusivo, este é *não* mutuamente exclusivo, o que implica que a ocorrência de um não exclui a possibilidade da ocorrência do outro. Por exemplo, é que possível a pessoa que atenda ao indivíduo seja uma mulher e também seja loira. Para a probabilidade de *eventos não mutuamente exclusivos* utilizamos a seguinte fórmula:

$$P(A \text{ ou } B) = P(A) + P(B) - P(A \text{ e } B) \quad (2.9)$$

Contudo, a probabilidade incógnita acima é:

$$P(A \text{ ou } B) = P(A) + P(B) - P(A \text{ e } B) = \frac{3}{10} + \frac{4}{10} - \frac{2}{10} = 0.50 = 50\% \quad (2.10)$$

2.3 Língua de Sinais

As línguas de sinais são idiomas gestuais humanos utilizados por comunidades de pessoas, geralmente surdas e/ou mudas, por todo o mundo, cujas formas consistem de seqüências de movimentos e configurações executados pelas partes do corpo, como mãos, braços e face. Há cerca de 40 anos, estudos começaram a mostrar que essas línguas são sistemas de idioma autênticos, com suas estruturas e regras, bem como os idiomas falados. Os pesquisadores utilizaram boa parte do tempo demonstrando as similaridades ocasionais entre as duas modalidades de linguagem: a falada e a sinalizada. As línguas de sinais são como línguas faladas, visto que elas exercem as mesmas funções sociais e mentais que os idiomas falados, e, além disto, podem ser aprendidas por crianças sem a necessidade de instruções, bastando apenas a exposição e interação normais com elas.

O sistema de linguagem é possível ao ser humano desde muitos milênios atrás por conta de sua evolução, e não existem comunidades que não utilizassem-no [20]. A língua de sinais nos mostra que o ser humano possui uma propensibilidade de se comunicar de duas maneiras, falada-ouvida e manual-visual. A primeira é predominante sobre a segunda, pois a porcentagem de pessoas bloqueadas de utilizar o meio de comunicação falado é muito menor que a de aptos. Com base em dados históricos, os primeiros relatos de comunidades utilizadoras de linguagens de sinais ocorreram há cerca de 300 anos atrás [20], o que justifica a sua menor popularização.

As línguas de sinais são linguagens naturais, ou seja, não foram inventadas conscientemente por uma pessoa. Elas se desenvolveram e se desenvolvem espontaneamente em comunidades de pessoas que têm a oportunidade de se comunicar regularmente umas com as outras. Elas possuem seus próprios vocabulários e estruturas gramaticais e não foram derivadas de línguas faladas¹, mas têm o mesmo poder de expressão que as linguagens faladas [20], porém, explorando meios físicos diferentes. Assim como os idiomas falados, existem vários idiomas gestuais diferentes no mundo, como a Língua Brasileira de Sinais e a Língua Americana de Sinais.

Como dito anteriormente, os gestos nas línguas de sinais não se limitam apenas a movimentos das mãos e braços, os chamados gestos manuais [13], mas também de sinais realizados por expressões faciais, movimentos da cabeça, do dorso e por posturas do corpo. Desta maneira, para o reconhecimento dos gestos de línguas de sinais é necessária uma observação simultânea dessas partes do corpo. Porém, neste trabalho, não foram consi-

¹Existem algumas línguas de sinais artificiais que são baseadas na língua falada, como a *Signed English*. No entanto, elas não são consideradas como sistemas de linguagem natural.

deradas todas estas características para a composição do gesto, mas apenas os movimentos das mãos, braço e cabeça.

Os gestos utilizados pelas línguas de sinais são geralmente considerados os mais estruturados dentre várias categorias de gestos [13]. Foram utilizados neste trabalho os gestos da Língua Brasileira de Sinais, que possui um conjunto de 46 diferentes configurações de mão [14]. Neste conjunto estão os gestos referentes às letras do alfabeto, sendo que 20 destas são representadas por posturas e 6 por gestos.

Posturas são sinais estáticos, ou seja, que não envolvem movimentação, assim como uma fotografia. Já os gestos são seqüências de posturas, que, conseqüentemente, formam um sinal com movimento, como as filmagens, que são seqüências de “fotografias”. Uma fotografia não possui movimento, ou seja, os objetos contidos nela estarão sempre estáticos na mesma posição. Por outro lado, em um vídeo os objetos geralmente se localizam em pontos diferentes com o decorrer do tempo, o que caracteriza a movimentação. Temos a sensação visual de movimentação nos vídeos porque eles são seqüências de fotografias tiradas em um determinado intervalo de tempo, e quando são exibidas seqüencialmente parecem estar em movimento. Com este mesmo raciocínio se caracteriza o gesto, sendo ele uma seqüência de posturas.



Figura 2.1: *Postura referente à letra “A” em LIBRAS.*

As Figuras 2.1 e 2.2 ilustram exemplos de postura e gesto, em que a primeira representa a postura para a letra “A” do alfabeto e a segunda o gesto para a letra “Y”.

Embora o alfabeto corresponda com o alfabeto da língua portuguesa, a língua de sinais não é apenas uma transcrição da língua falada para a gesticulada. Apenas em casos especiais são utilizadas seqüências de símbolos, como os do alfabeto, para formação de palavras, porém, eles aparecem como parte integrante em outros sinais. Por exemplo, a palavra “difundir” é repre-

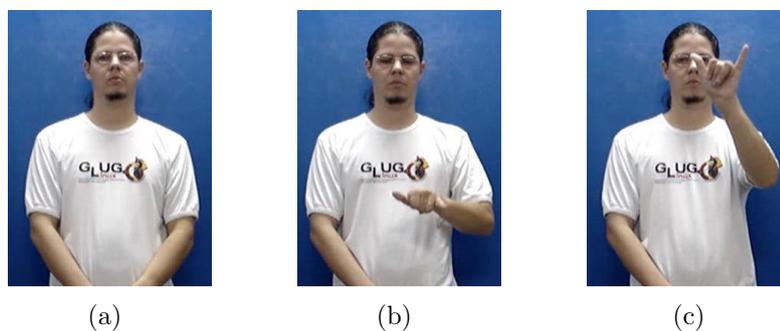


Figura 2.2: *Imagens do gesto referente à letra “Y” em LIBRAS.*

sentada pelas mãos em configuração de “Y”² tocando o queixo e, em seguida, movendo-se para os lados opostos, abrindo os braços.

²A configuração de mão referente à letra “Y” pode ser observada na Figura 2.2(c)

Capítulo 3

Modelo de Markov Oculto

Durante os últimos 15 anos, os modelos de Markov ocultos têm sido amplamente utilizados em diversas áreas, incluindo no reconhecimento de padrões [10]. Dentre elas podem ser destacadas as aplicações de reconhecimento de voz, em que predominam a utilização de modelos estatísticos, em destaque os baseados em modelo de Markov oculto [25].

Antes de descrever o modelo de Markov oculto (*HMM - Hidden Markov Model*), é necessário conhecer sobre sua origem, os processos de Markov. Em meados de 1907, Markov definiu e investigou algumas propriedades que hoje são conhecidas como processos de Markov [2]. A principal característica dos processos de Markov é a definição que toda a história passada está resumida no valor atual do processo. Em alguns padrões geralmente existe uma estrutura que influencia a probabilidade do próximo evento ocorrer. Por exemplo, em alguns idiomas, como no Português e no Inglês, a probabilidade de se encontrar a letra u após ter detectado a letra q é muito alta, considerando que praticamente sempre após a letra q é encontrada a letra u . Um processo estocástico é chamado de processo de Markov de ordem j se a probabilidade do evento corrente ocorrer, dados todos os eventos anteriores e o presente, depende somente dos j eventos mais recentes.

Um modelo de Markov oculto é uma Cadeia de Markov onde os estados do modelo não são conhecidos, mas apenas o sinal emitido em cada unidade de tempo t . Existem duas camadas estocásticas nos modelos de Markov ocultos. A primeira camada estocástica é um processo de Markov de primeira ordem, porém, não é diretamente observável. Em um processo de Markov cada estado é uma possível observação, e existe uma probabilidade de transição de um determinado estado para outro, ou seja, é a probabilidade do estado corrente ser S_y no tempo t , estando no estado S_x no tempo $t-1$ ¹. A segunda

¹O termo “observações” recebe diferentes significados quando aplicados a processos de

camada estocástica é um conjunto de probabilidades para cada estado do modelo que indicam a probabilidade de uma observação ocorrer estando em determinado estado. A seqüência de estados percorrida em um modelo, dada a seqüência de observações, é oculta ao observador. Ou seja, dada uma seqüência de saída, não se sabe a seqüência de estados percorrida pelo modelo, mas somente uma função probabilística deste caminho, e por isso o modelo é chamado de *modelo de Markov oculto* ou *Hidden Markov Model (HMM)*.

3.1 Elementos do HMM

Um modelo de Markov oculto geralmente é definido por uma tripla $\lambda = (A, B, \pi)$, e possui os seguintes elementos:

1. Número de estados do modelo, N , sendo que o conjunto de estados existentes é denotado por $S = \{s_1, s_2, \dots, s_N\}$.
2. O número total de símbolos distintos reconhecidos, M , em que $V = \{v_1, v_2, \dots, v_M\}$ representa o alfabeto de sinais do modelo.
3. A quantidade de símbolos existentes na seqüência de observação, T , ou seja, o tamanho da seqüência de observação.
4. Um conjunto de estados $Q = \{q_1, q_2, \dots, q_T\}$ para determinada observação, em que q_t é o estado no momento t da observação.
5. Um conjunto de símbolos $O = \{o_1, o_2, \dots, o_T\}$ para determinada observação, em que o_t é o símbolo observado no instante t .
6. Um conjunto de valores, π , que define as probabilidades de cada estado ser um estado inicial, $\pi = \{\pi_1, \pi_2, \dots, \pi_N\}$.
7. Uma matriz bidimensional $A_{N \times N}$, que possui os valores das probabilidades de transições entre os estados do modelo.
8. Uma matriz bidimensional $B_{N \times M}$, que representa as respectivas probabilidades de ocorrências dos símbolos em cada estado do modelo.

Supondo um HMM onde existam 3 estados, s_1 , s_2 e s_3 , e as probabilidades de uma cadeia iniciar no estado s_1 é de 25%, no s_2 35%, e no s_3 40%. Neste

Markov e a modelos de Markov ocultos. No primeiro, essas observações são os estados percorridos pelo modelo, enquanto que nos modelos de Markov ocultos são os símbolos gerados nos estados do modelo.

modelo existem três símbolos possíveis que podem ser gerados em cada estado do modelo: $V = \{v_1, v_2 \text{ e } v_3\}$, e a probabilidade de, estando no estado s_1 , ocorrer um v_1 é de 7%; de ocorrer um v_2 é de 63%; e um v_3 é 30%. Estando em s_2 as probabilidades serão de² 30% para ocorrer um v_1 e 70% de ocorrer um v_3 . Por fim, estando em s_3 as probabilidades serão de 83% de ocorrer um v_1 , 3% de ocorrer um v_2 e 14% de ocorrer um v_3 . Estando no estado s_1 , existe 28% de chance que ele permaneça em s_1 no próximo instante de tempo ($t + 1$), 53% de chance de ir para o estado s_2 , e 19% de chance de ir para s_3 ; estando em s_2 é impossível retornar a ele, e pode tanto ir para s_1 quanto para s_3 com a mesma probabilidade; finalmente, estando em s_3 existe 14% de possibilidade que ele vá para o estado s_1 , 85% para s_2 e 1% que fique aonde está. Com base neste modelo foi observada a seguinte seqüência de observação: v_3, v_2, v_2, v_1 e v_3 .

Com base nestas informações, os elementos do modelo são configurados com os valores abaixo, e as probabilidades para a seqüência de observações podem ser visualizadas na Figura 3.1. No gráfico, o quadrado à esquerda representa o início de qualquer seqüência possível, pois está ligado às probabilidades de ocorrência inicial dos estados. Os círculos e as linhas podem ser analogamente comparados aos vértices e arestas dos grafos, respectivamente. Ligados às arestas existem valores numéricos, que indicam as probabilidades de transição do estado mais à esquerda para o outro estado ligado a ele. Cada linha horizontal representa um estado, sendo, neste caso, três estados. O número de vértices em cada uma dessas linhas é relacionado com o tamanho da seqüência de símbolos observada. Em cada vértice é exibida a probabilidade de gerar o símbolo observado em cada unidade de tempo em cada estado.

$$N = 3 \text{ (referente a } s_1, s_2 \text{ e } s_3\text{);}$$

$$M = 3 \text{ (referente aos símbolos } v_1, v_2 \text{ e } v_3\text{)}$$

$$A_{N \times N} = \left\{ \begin{array}{ccc} 0.28 & 0.53 & 0.19 \\ 0.50 & 0.00 & 0.50 \\ 0.14 & 0.85 & 0.01 \end{array} \right\}$$

$$B_{N \times M} = \left\{ \begin{array}{ccc} 0.07 & 0.63 & 0.30 \\ 0.30 & 0.00 & 0.70 \\ 0.83 & 0.03 & 0.14 \end{array} \right\}$$

$$\pi_N = [0.25 \quad 0.35 \quad 0.40]$$

²Note que no estado s_2 a probabilidade de ocorrer um v_2 é nula.

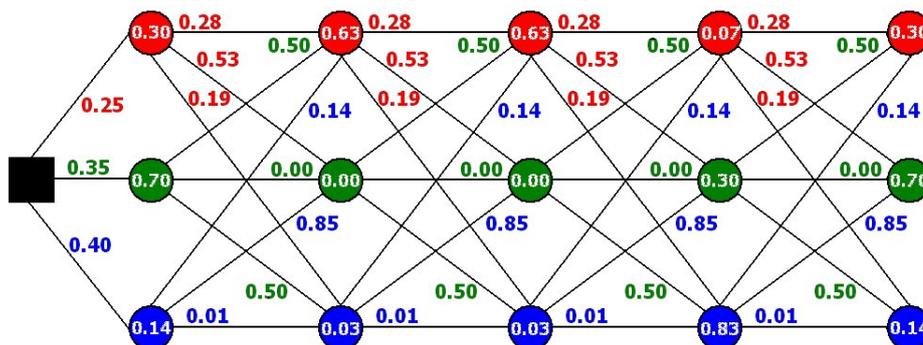


Figura 3.1: Treliça que representa o exemplo de modelo de Markov oculto e um possível observação.

3.2 Problemas Básicos

Existem três problemas principais implícitos no HMM, e as suas soluções contribuem para o funcionamento eficaz das aplicações do mundo real que o utilizam [10]. Antes de descrever o problema, é interessante saber que sempre que no decorrer do texto forem encontrados os símbolos λ , eles se referem a HMMs; a letra O a uma seqüência de símbolos observados; e $P[O|\lambda]$ é a probabilidade da seqüência de observação O ter sido gerada pelo modelo λ .

3.2.1 Problema de avaliação

O primeiro problema se refere à descoberta da probabilidade de uma determinada seqüência de observações O ter sido gerada por um específico modelo λ . Este tipo de situação pode ser muito freqüente nas aplicações de HMM. Como exemplo, no reconhecimento de voz, ao se produzir um fonema qualquer, esta entrada pode ser classificada como pertencente ao modelo que indicar a maior probabilidade ($P[O|\lambda]$); outro exemplo seria no reconhecimento de comportamento animal, em que o modelo que apresentar a maior probabilidade indica a reação do animal perante uma situação ou um conjunto de estímulos; o mesmo se aplica em um sistema reconhecedor de palavras manuscritas, em que para cada palavra exista um modelo e o modelo com a maior probabilidade de ter gerado a observação classifica a palavra.

A maneira mais simples de calcular a probabilidade de determinada seqüência de observação ter sido gerada por um modelo é através da verificação de todas as seqüências de estados de tamanho T (número total de observações) possíveis, e posteriormente calcular suas probabilidades. Con-

siderando uma destas seqüências como $Q = \{q_1, q_2, \dots, q_T\}$, em que q_1 é o estado inicial, a probabilidade de uma seqüência de símbolos ter sido gerada por esta seqüência de estados Q é a seguinte:

$$P(O|Q, \lambda) = \prod_{i=1}^T b_i(o_i) \quad (3.1)$$

em que $b_i(o_i)$ é a probabilidade de se gerar um símbolo o_i estando no estado q_j . A probabilidade da seqüência de estados Q ocorrer pode ser representada pela seguinte equação:

$$P(Q|\lambda) = \pi_1 \cdot a_{1,2} \cdot a_{2,3} \dots a_{T-1,T} \quad (3.2)$$

em que π_i é a probabilidade do estado inicial do modelo ser q_i e $a_{i,j}$, a probabilidade de transição do estado q_i para q_j .

Por fim, a probabilidade de O ter sido gerada pelo modelo é a soma das duas probabilidades acima para todas as seqüências de estados de Q , da seguinte forma:

$$P(O|\lambda) = \sum_{allQ} P(O|Q, \lambda) \cdot P(Q|\lambda) \quad (3.3)$$

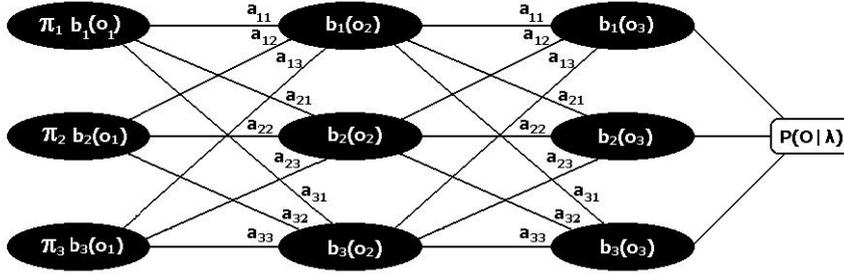
A probabilidade $P(O|\lambda)$ através das equações acima é obtida com cálculos de ordem computacional $2T \cdot N^T$, pois existem N possíveis estados que podem ser alcançados para cada $t = 1, 2, \dots, T$, ou seja, existem N^T possíveis seqüências de estados, e para cada seqüência desta $2 \cdot T$ cálculos. Estes cálculos são computacionalmente impraticáveis, até mesmo quando os valores de N e T são baixos. Obviamente, é necessário um procedimento mais eficiente para resolver este problema. Este procedimento existe, e é chamado de procedimento *forward* [17].

Para a representação deste procedimento utiliza-se a variável α , sendo referente à probabilidade do estado corrente no tempo t ser q_i através da seqüência de observações $O = \{o_1, o_2, \dots, o_t\}$ (até o tempo t), em um dado λ . É possível executar o cálculo desta variável α recursivamente, através dos seguintes passos:

1. Inicialização

$$\alpha_1(i) = \pi_i \cdot b_i(o_1), 1 \leq i \leq N \quad (3.4)$$

2. Indução

Figura 3.2: Treliça referente ao algoritmo *forward*.

$$\alpha_{t+1}(j) = \left[\sum_{i=1}^N \alpha_t(i) \cdot a_{ij} \right] \cdot b_j(o_{t+1}), 1 \leq t \leq T-1 \text{ e } 1 \leq j \leq N \quad (3.5)$$

3. Terminação

$$P(O|\lambda) = \sum_{i=1}^N \alpha_T(i) \quad (3.6)$$

Para melhor compreensão do procedimento *forward* e sua melhora em desempenho, vamos estudar o seguinte exemplo. Considere que temos uma seqüência de observação O de tamanho 3 e um modelo de Markov oculto λ constituído por 3 estados, M símbolos, as respectivas matrizes de probabilidades A e B e o conjunto π . A treliça mostrada na Figura 3.2 representa o nosso exemplo, sendo que cada coluna de vértices representa uma unidade de tempo na seqüência de símbolos observada e cada linha cada estado do modelo. A intersecção entre o número x horizontal e y vertical chega a um vértice, demonstrado por um círculo, que representa a probabilidade de estar no estado s_y e no instante x da seqüência de observação.

A probabilidade de se observar o primeiro símbolo da seqüência de observação estando no estado s_1 de nosso exemplo é a probabilidade conjunta deste estado ser inicial (π_1) e gerar o primeiro símbolo da observação ($b_1(o_1)$), portanto, $\pi_1 \cdot b_1(o_1)$. O mesmo se aplica para os estados 2 e 3, conforme demonstrado nos vértices da coluna 1 da treliça exibida na Figura 3.2. Já a probabilidade de se observar o segundo símbolo da seqüência de observação estando no estado s_x , como são eventos independentes, é a multiplicação da probabilidade do estado s_x gerar o símbolo o_2 ($b_x(o_2)$) e o somatório das probabilidades de estar em cada estado do modelo no momento anterior, gerar o

símbolo o_1 da seqüência O e transitar deles para s_x . Como a primeira linha da treliça é correspondente às probabilidades de gerar o primeiro símbolo em cada um dos estados, o valor do somatório já está praticamente calculado. Foi multiplicado o valor da probabilidade do estado gerador do primeiro símbolo ser s_1 com a probabilidade de transição de s_1 para s_x e a probabilidade de s_x gerar o_2 e temos um dos valores do somatório, conseqüentemente, executando o mesmo processo para os outros estados obtemos o somatório completo.

A cada um dos cálculos de probabilidade, o algoritmo *forward* armazena estas probabilidades em suas variáveis. Por exemplo, a probabilidade de ter gerado o símbolo o_1 no estado s_3 estará atribuída à variável $\alpha_1(3)$, e assim sucessivamente. Desta maneira, evitam-se cálculos redundantes e, conseqüentemente, ganha-se em desempenho. Esse ganho pode ser observado no cálculo da probabilidade de gerar o símbolo o_2 no estado s_2 ($\alpha_2(2)$), seguindo a ordem de execução do processo *forward*, em que não será necessário calcular os vértices da primeira coluna novamente, pois eles já foram encontrados durante o cálculo da probabilidade de gerar o símbolo o_2 no estado s_1 . Em nosso exemplo, o cálculo de $\alpha_3(1)$ e $\alpha_3(2)$ pelo processo manual resulta em 60 multiplicações e 16 somas, e utilizando o processo *forward* 23 multiplicações e 10 somas.

Alternativamente, existe outra maneira de se obter a probabilidade desejada, que é através do procedimento *backward*, representado pela variável β . O algoritmo *backward*, assim como o *forward*, utiliza a estratégia de programação dinâmica para minizar o custo computacional do cálculo. Contudo, vale lembrar que os procedimentos *forward* e *backward* são independentes e apenas uma das variáveis, α ou β , é necessária para solucionar o problema da avaliação. A variável $\beta_t(i)$ é a probabilidade da seqüência de observação de $t+1$ até o fim, dado o estado S_i no tempo t e o modelo λ . Indutivamente, β pode ser calculada da seguinte maneira:

1. Inicialização

$$\beta_T(i) = 1, 1 \leq i \leq N. \quad (3.7)$$

2. Indução

$$\beta_t(i) = \sum_{j=1}^N a_{ij} \cdot b_j(o_{t+1}) \cdot \beta_{t+1}(j), t = T-1, T-2, \dots, 1, 1 \leq i \leq N. \quad (3.8)$$

3. Terminação

$$P(O|\lambda) = \sum_{i=1}^N \pi_i \cdot b_i(o_1) \cdot \beta_1(i) \quad (3.9)$$

3.2.2 Problema da busca da melhor seqüência de estados

A solução do segundo problema busca encontrar a melhor seqüência de estados percorrida pelo modelo para uma determinada seqüência de observação. Ou seja, a seqüência cuja probabilidade final seja a maior dentre as outras. Geralmente ele é resolvido utilizando o algoritmo de Viterbi [10], que localiza esta melhor seqüência de estados $Q = \{q_1, q_2, \dots, q_T\}$ para a seqüência de observações $O = \{o_1, o_2, \dots, o_T\}$ dada. Este problema e sua possível solução não serão discutidos neste trabalho, pois não foram utilizados nesta aplicação. Porém, uma possível utilização dessa solução, direcionada a este tema, seria a análise das seqüências de estados mais prováveis de cada modelo para determinadas observações de símbolos, a fim de se encontrar padrões que auxiliem no processo de classificação, como a ausência de transição entre estados, por exemplo.

3.2.3 Problema de treinamento

O terceiro e último problema é também o mais difícil de solucionar, que é encontrar um método para estimar os parâmetros $\lambda(A, B, \pi)$ do modelo de forma que atenda a um critério de otimização. A seqüência O utilizada para este ajuste é chamada de seqüência de treinamento, pois é utilizada para treinar o HMM. Neste trabalho, para ajustar os parâmetros dos modelos construídos foram utilizados valores estatísticos obtidos dos resultados de análises visuais executadas em vídeos presentes em um banco de imagens, que será descrito posteriormente.

Não existe uma maneira conhecida de realizar este ajuste para resolver analiticamente o modelo que maximize a probabilidade da seqüência de observações, porém, é possível escolher o modelo que sua probabilidade seja localmente maximizada usando um procedimento iterativo. A existência de um processo que ajusta iterativamente e automaticamente os parâmetros do modelo com base nas seqüências de observações é a principal força dos HMMs [6], que é o algoritmo de Baum-Welch [17], que se trata de uma especialização do algoritmo EM - *Expectation-Maximization* [3, 16] aplicada aos HMMs.

Antes de descrever o método de reestimação, é interessante que sejam

definidas algumas variáveis que serão utilizadas ao longo do processo. Primeiramente, define-se a variável $\xi_t(i, j)$ como sendo a probabilidade de estar no estado s_i no momento t e transitar para o estado s_j no momento seguinte, dado o modelo e a observação. Desta forma, a Equação 3.10 descreve $\xi_t(i, j)$ matematicamente, considerando $Q = q_1, q_2, \dots, q_T$ a seqüência de estados e $S = s_1, s_2, \dots, s_N$ o conjunto de estados do modelo

$$\xi_t(i, j) = P(q_t = s_i, q_{t+1} = s_j | O, \lambda), \quad (3.10)$$

Dado que $\alpha_t(i)$ é a probabilidade de estar no estado s_i no tempo t desde o início da observação e que $\beta_t(i)$ é a probabilidade de geração da seqüência no modelo do instante $t + 1$ ao fim, estando no estado s_i no tempo t , então $\xi_t(i, j)$ pode ser reescrito com o auxílio das variáveis α e β , descritas na Seção 3.2.1, da seguinte forma:

$$\begin{aligned} \xi_t(i, j) &= \frac{\alpha_t(i) \cdot a_{ij} \cdot b_j(o_{t+1}) \cdot \beta_{t+1}(j)}{P(O|\lambda)} = \\ &= \frac{\alpha_t(i) \cdot a_{ij} \cdot b_j(O_{t+1}) \cdot \beta_{t+1}(j)}{\sum_{i=1}^N \sum_{j=1}^N \alpha_t(i) \cdot a_{ij} \cdot b_j(o_{t+1}) \cdot \beta_{t+1}(j)} \end{aligned} \quad (3.11)$$

Outra variável a ser compreendida é $\gamma_t(i)$, que representa a probabilidade de estar no estado s_i no instante de tempo t , como descreve a equação abaixo:

$$\gamma_t(i) = P(q_t = s_i | O, \lambda) \quad (3.12)$$

Da mesma forma, deduz-se $\gamma_t(i)$ através de α e β da seguinte maneira:

$$\gamma_t(i) = \frac{\alpha_t(i) \cdot \beta_t(i)}{\sum_{i=1}^N \alpha_t(i) \cdot \beta_t(i)} \quad (3.13)$$

Pode-se ainda relacionar ξ com γ da seguinte forma:

$$\gamma_t(i) = \sum_{j=1}^N \xi_t(i, j) \quad (3.14)$$

Nesse sentido, observa-se que a probabilidade de partir do estado s_i para s_j na seqüência de observação é o somatório das probabilidades de estar em cada instante da observação no estado s_i e transitar para ao estado s_j . Ou seja,

$$\sum_{t=1}^{T-1} \xi_t(i, j) \quad (3.15)$$

Já a probabilidade de estar no estado s_i e partir para um estado qualquer, pode ser escrita através da Equação 3.16:

$$\sum_{t=1}^{T-1} \gamma_t(i) \quad (3.16)$$

Com as definições acima temos os seguintes resultados:

$$\pi_i = \text{probabilidade de estar no estado } s_i \text{ no instante } 1 = \gamma_1(i) \quad (3.17)$$

$$a_{ij} = \frac{\text{número de transições do estado } s_i \text{ para o estado } s_j}{\text{número de transições do estado } s_i \text{ para qualquer estado}} \quad (3.18)$$

$$a_{ij} = \frac{\sum_{t=1}^{T-1} \xi_{t,i,j}}{\sum_{t=1}^{T-1} \xi_{t,i}} \quad (3.19)$$

$$b_i(k) = \frac{\text{número de vezes que foi observado o símbolo } o_k \text{ no estado } s_i}{\text{número de vezes no estado } s_i} \quad (3.20)$$

$$b_i(k) = \frac{o_{t=v_k} \sum_{t=1}^T \gamma_t(i)}{\sum_{t=1}^T \gamma_t(i)} \quad (3.21)$$

Dessa forma, temos todas as incógnitas para a reestimação do modelo a partir da observação. Portanto, o algoritmo Baum-Welch realiza diversas iterações sobre a seqüência de observações dada, com o intuito de reestimar as componentes do HMM com base nessa entrada. Deve ser estabelecido um critério de parada para o algoritmo, como a estabilização dos valores das componentes de uma iteração para a outra.

Capítulo 4

Desenvolvimento

4.1 Modelagem, Captura e Análise dos Gestos

A Língua Brasileira de Sinais, LIBRAS, possui uma extensa variedade de gestos utilizados na formação de suas sentenças. Dentre eles, foram selecionados gestos provenientes do dicionário trilíngüe [18, 19]. Para executar a seleção, foram utilizados como critérios a complexidade do gesto, ou melhor, a complexidade do movimento do gesto, o nível de confusão que ele pode gerar com outros gestos já escolhidos, devido a semelhança entre eles, e a representatividade do gesto. Foram escolhidos gestos com movimentos complexos com intuito de testar a eficiência da aplicação da técnica, que será utilizada em ambientes em que o usuário movimentava seus braços, mãos e cabeça; o segundo critério foi utilizado para analisar a atuação do classificador diante de gestos semelhantes; e o terceiro para explorar de maneira enfática algumas configurações de postura do corpo, como a inflação das bochechas, por exemplo.

Para cada gesto foi desenvolvido um respectivo modelo de Markov oculto. Para a construção desses modelos foram escolhidas as posturas mais marcantes dos gestos, sendo que cada uma destas posturas se relaciona diretamente a um estado do modelo. Por exemplo, a Figura 4.1 ilustra as posturas marcantes para execução do gesto referente a “derrame-cerebral”, em que a mão direita deve estar em M, com a palma virada para a esquerda e com os dedos apontando para cima, tocando o lado direito da testa (Figura 4.1(a)). Em seguida, cruzar os braços em frente à cabeça, com as duas mãos fechadas e com as palmas para dentro (Figura 4.1(b)), e mover as mãos para os lados opostos com força (Figura 4.1(c)). Como foram consideradas três posturas como marcantes, conseqüentemente, o modelo construído para a

representação desta expressão conterà três estados.

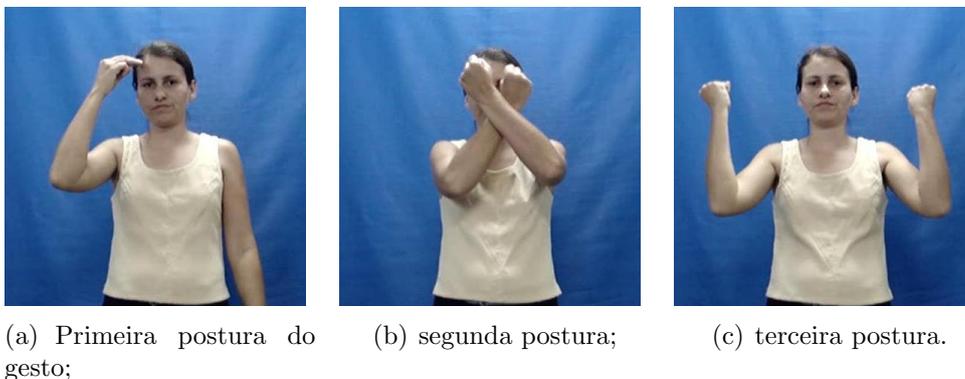


Figura 4.1: *Posturas do gesto referente à expressão “derrame cerebral” em LIBRAS.*

Como apenas as posturas denominadas como marcantes têm um respectivo estado no modelo, algumas posturas são “ignoradas”¹, predominando apenas as mais importantes em nosso conceito. Uma postura é uma configuração estática, sem movimento, enquanto o gesto é dinâmico, ou seja, possui movimento. Por exemplo, a foto de uma mão e a filmagem de uma cabeça se deslocando da esquerda para a direita são exemplos de postura e gesto, respectivamente. De uma postura o usuário transita para outra postura, que, conseqüentemente, produz um gesto, e, com isto, obtêm-se as transições de estados do modelo. No entanto, um modelo de Markov oculto não se constitui apenas de estados e transições, mas também necessita das matrizes de probabilidades de transição de estados e de geração de símbolos, além do conjunto de probabilidades iniciais.

Primeiramente, foram definidos as características que, juntas, nos auxiliam a discriminar os gestos executados pelos usuários. As características foram escolhidas com base nas descrições e padronizações contidas em [18, 19] e na análise dos gestos escolhidos, observando quais combinações discriminavam os gestos selecionados de maneira única. Essas características são as seguintes: i) posição espacial vertical da mão (PEV); ii) posição espacial horizontal da mão (PEH); iii) configuração da mão (CON); iv) orientação da mão (ORI); v) direção da palma da mão (DIP); e vi) situação das bochechas (SIB). Neste sentido, cada postura do gesto é configurada com a combinação destas características, sendo que todas as que se referem a in-

¹As posturas contidas no gesto que são classificadas como desconhecidas para o modelo são aproximadas pelo sistema a uma das posturas marcantes, obedecendo a determinado critério de classificação.

formações da mão (PEV, PEH, CON, ORI e DIP) se transformam em duas características: mão direita e mão esquerda. Portanto, cada postura se classifica pelo conjunto de 11 características. Porém, este conjunto apresenta algumas limitações, caso o conjunto de gestos seja estendido, pois foram extraídas as características necessárias para distinguir apenas os gestos selecionados. Portanto, características como expressão facial e configuração dos ombros não foram consideradas como atributos relevantes para o reconhecimento de cada gesto.

Para melhor compreensão e visualização destas características na imagem, considere as Figuras 4.2 e 4.3. As posições espaciais vertical e horizontal, como o próprio nome diz, são as localizações das mãos em relação a um eixo imaginário que corta o centro do corpo do usuário.

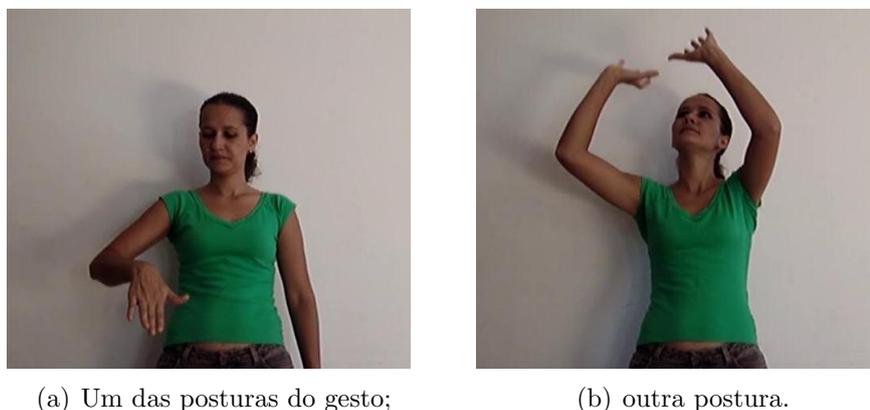


Figura 4.2: *Algumas posturas do gesto referente à palavra “basquete” em LIBRAS.*

A posição espacial vertical (PEV) pode ser observada nas Figuras 4.2(a) e 4.2(b), pois na primeira a mão direita do usuário está localizada entre a cintura e o peito, enquanto que na segunda está acima da cabeça. A posição horizontal (PEH) também pode ser visualizada na Figura 4.3, em que as mãos estão primeiramente ao lado direito do corpo e, em seguida, partem para o lado esquerdo.

A configuração das mãos (CON) pode ser observada na Figura 4.4, em que o usuário está com a mão na configuração da letra “L” na Figura 4.4(a), enquanto que em outro instante do gesto, representado pela Figura 4.4(b), a mão está na configuração da letra “S”².

²A letra “L”, em LIBRAS, se caracteriza pela mão fechada e os dedos indicador e polegar distendidos, e a letra “S” pela mão fechada, com o polegar cobrindo os outros dedos.

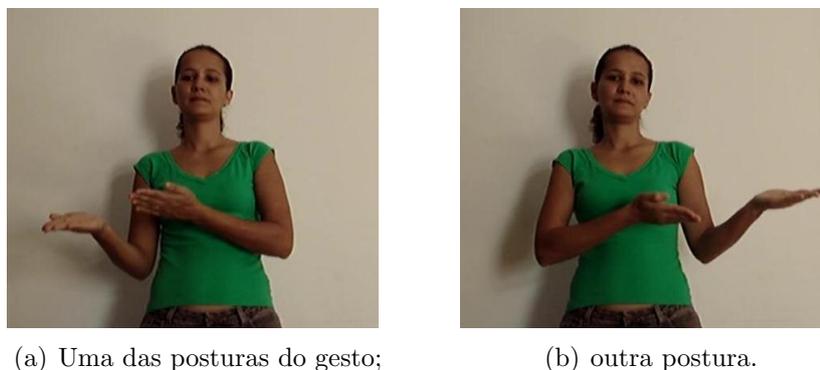


Figura 4.3: *Algumas posturas do gesto referente à palavra “bem-vindo” em LIBRAS.*

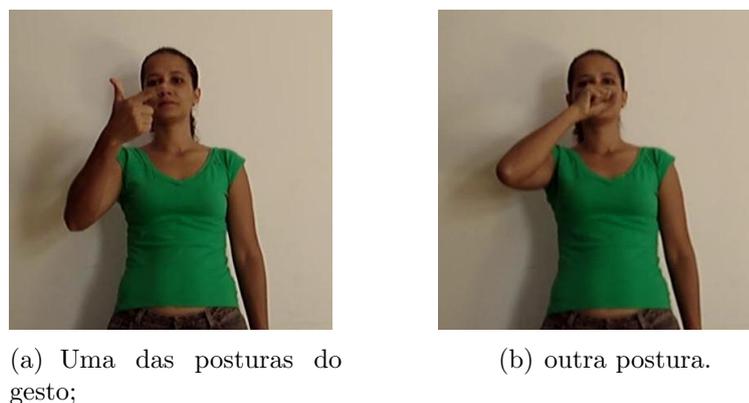


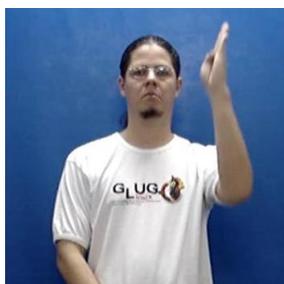
Figura 4.4: *Algumas posturas do gesto referente à palavra “inodoro” em LIBRAS.*

A orientação (ORI) da mão e a direção da palma da mão (DIP) também podem ser observadas na Figura 4.5, referente ao sinal “mau” em LIBRAS³, sendo nas Figuras 4.5(a) a mão esquerda na posição vertical com a palma para a direita e em 4.5(b) na posição horizontal com a palma para baixo.

E, por fim, a situação das bochechas pode ser facilmente observada pelas Figuras 4.6(a) e 4.6(b), em que na segunda o usuário está com as bochechas infladas, o que não ocorre na primeira.

Como foram utilizados modelos discretos, as características acima mencionadas devem receber apenas valores entre intervalos conhecidos, ou seja, em nosso caso todos os valores possíveis para cada característica são previa-

³A imagem do gesto referente à palavra “mau” em LIBRAS não confere com a descrição do dicionário utilizado, pois o usuário exibido é canhoto.



(a) Primeira postura do gesto;



(b) segunda postura.

Figura 4.5: *Posturas do gesto referente à palavra “mau” em LIBRAS, executadas por um usuário canhoto.*



(a) Primeira postura do gesto;



(b) segunda postura.

Figura 4.6: *Posturas do gesto referente à palavra “demolir” em LIBRAS.*

mente conhecidos. A Tabela 4.1 descreve as características e as atribuições aceitas para cada uma delas.

Com base nesta padronização, os vídeos referentes aos gestos foram classificados um a um e quadro a quadro. Como um vídeo é composto de uma seqüência de imagens (quadros), foram extraídas as características acima para todas as imagens de cada vídeo, e, posteriormente, foram atribuídos valores a elas, através de análise visual. As informações oriundas do processo anterior foram úteis para a formação das matrizes de probabilidades (matrizes de ocorrências iniciais e transição de estados e matriz de geração de símbolos) de cada modelo de Markov oculto.

Em nossos modelos, como dito anteriormente, cada postura marcante do gesto se relaciona a um estado, e cada postura contida na imagem, sendo mar-

PEV	entre a cintura e o peito; entre o peito e os ombros; na altura do queixo; na altura da boca; na altura do nariz; na altura dos olhos; acima dos olhos e abaixo do topo da cabeça; acima da cabeça; e desconhecida.
PEH	à direita, muito afastada do corpo; à direita, afastada do corpo; à direita, próxima ao corpo; ao centro do corpo; à esquerda, próxima ao corpo; à esquerda, afastada do corpo; à esquerda, muito afastada do corpo; desconhecida.
CON	mão em A; mão em S; mão em L; mão em 1; mão em D; mão fechada com o dedo polegar distendido; mão fechada com dedos indicador e polegar distendidos tocando-se; mão fechada com dedos indicador e polegar distendidos em paralelo; mão fechada com todos os dedos distendidos tocando-se; mão em V; mão em U; mão em M; mão em Y; mão em 5; mão em 2, mão em B; mão em C; mão aberta com os dedos unidos; mão aberta com os dedos separados; mão aberta com os dedos unidos e curvados; mão aberta com os dedos separados e curvados; desconhecida.
ORI	horizontal; vertical; desconhecida.
DIP	para frente; para trás; para a direita; para a esquerda; para cima; para baixo; desconhecida.
SIB	ambas infladas; ambas não infladas.

Tabela 4.1: Conjunto de atributos caracterizadores de postura e seus respectivos valores permitidos.

cante ou não, produzirá uma observação do modelo. Desta forma, em cada estado do modelo temos as probabilidades de cada configuração de valores dos atributos, ou seja, temos as probabilidades dos valores das características do conjunto. A Figura 4.7 ilustra as três posturas que denominamos marcantes no gesto referente à palavra “abafado” em LIBRAS.

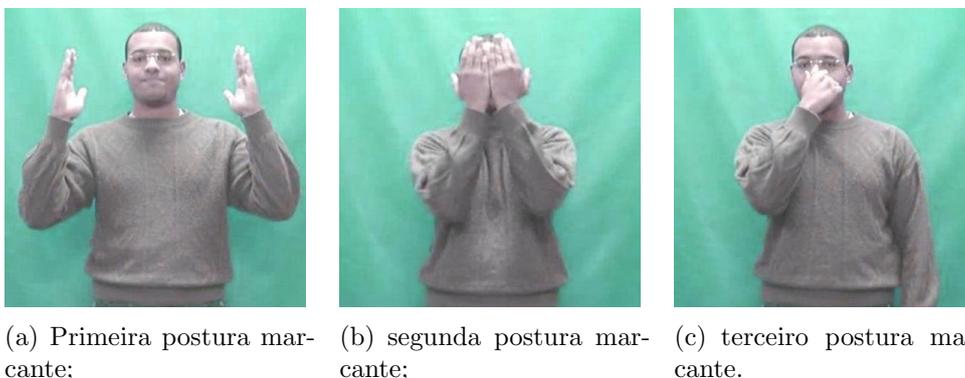


Figura 4.7: *Posturas do gesto referente à palavra “abafado” em LIBRAS que foram consideradas como marcantes.*

Na Figura 4.7(a), como se pode observar, o usuário está com as mãos verticais abertas, palma a palma, e com os dedos ligeiramente flexionados a cada lado da cabeça. Em continuação ao gesto, o usuário move as mãos para diante do rosto até que elas se toquem pelos dedos mínimos, alcançando a postura ilustrada pela Figura 4.7(b). Finalizando o gesto, o usuário faz o sinal de A com a mão direita, com a palma para dentro, e aperta o nariz com o polegar e o indicador, assim como mostra a Figura 4.7(c). A Tabela 4.2 exibe os valores de todas as características para cada postura do gesto, em que os símbolos D e E representam determinada característica para a mão direita e para a esquerda, respectivamente.

A Figura 4.8 simboliza uma possível amostra do gesto “abafado”. Como dito anteriormente, algumas posturas do gesto, por não serem especificamente as escolhidas como marcantes, devem ser enquadradas como uma das marcantes. Esse fato pode ser observado nas Figuras 4.8(a), 4.8(b), 4.8(c), em que para cada postura foi atribuído o estado mais adequado, denominado como ABF1, ABF2 e ABF3, que se relacionam diretamente com as posturas marcantes citadas na Figura 4.7.

Como demonstrado na situação anterior, as combinações das características com seus respectivos valores formam os estados de cada modelo, e em cada estado existem as probabilidades de geração para cada símbolo reconhecido pelo modelo. Por exemplo, o conjunto dos valores das características exibidos na Tabela 4.1 seriam os três possíveis estados do modelo que descreve o gesto “abafado”, e para cada uma destas observações existiriam as probabilidades para as características extraídas da imagem. Dentre os benefícios do HMM está o fato de que os modelos de gestos podem ser treinados automaticamente com base em séries de exemplos de classes de



(a) Posturas consideradas como pertencentes ao estado ABF1 durante a análise visual.



(b) Posturas consideradas como pertencentes ao estado ABF2 durante a análise visual.



(c) Posturas consideradas como pertencentes ao estado ABF3 durante a análise visual.

Figura 4.8: *Seqüência de posturas referentes ao gesto “abafado”, em LIBRAS.*

gestos, com isto, o modelo treinado codifica a variação existente no conjunto de exemplos [24]. Desta forma, essas probabilidades serão obtidas através da análise e classificação manual dos vídeos de treinamento do sistema.

Os gestos selecionados foram extraídos do dicionário LIBRAS [18, 19], em que foram escolhidos os gestos que possuíam uma movimentação acentuada e os que poderiam ser confundidos com outros previamente selecionados. As descrições dos gestos foram retiradas do dicionário acima citado, porém,

PEV-D	1) na altura dos olhos
	2) na altura dos olhos
	3) na altura da boca
PEV-E	1) na altura dos olhos
	2) na altura dos olhos
	3) desconhecida
PEH-D	1) à direita, afastada do corpo
	2) ao centro do corpo
	3) ao centro do corpo
PEH-E	1) à esquerda, afastada do corpo
	2) ao centro do corpo
	3) desconhecida
CON-D	1) mão aberta com os dedos unidos e curvados
	2) mão aberta com os dedos unidos e curvados
	3) mão em A.
CON-E	1) mão aberta com os dedos unidos e curvados
	2) mão aberta com os dedos unidos e curvados
	3) desconhecida
ORI-D	1) vertical
	2) vertical
	3) vertical
ORI-E	1) vertical
	2) vertical
	3) desconhecida
DIP-D	1) para a esquerda
	2) para trás
	3) para trás
DIP-E	1) para a direita
	2) para trás
	3) desconhecida
SIB	1) ambas não infladas
	2) ambas não infladas
	3) ambas não infladas

Tabela 4.2: Uma das possíveis configurações que podem surgir da imagem do usuário executando as posturas ilustradas pela Figura 4.7, sendo que os valores seguidos do número 1 correspondem à Figura 4.7(a), 2 à 4.7(b) e 3 à 4.7(c).

alguns gestos foram modificados, conseqüentemente, suas descrições também foram alteradas em nosso documento. Como exemplo, o gesto referente à palavra “babá” no dicionário inicia com a mão direita em 1, palma para dentro, tocando a ponta do indicador abaixo do olho direito. Porém, esta movimentação inicial não existe no dialeto LIBRAS de Mato Grosso do Sul, portanto, foi alterada sua descrição.

Antes de descrever esses gestos iremos padronizar os valores possíveis para as características (PEV, PEH, CON, ORI, DIP e SIB) das posturas. Cada valor receberá um respectivo número seqüencial para representá-lo, começando de 1 (um). Como exemplo, para a característica PEV temos: entre a cintura e o peito = 1; entre o peito e os ombros = 2; na altura do queixo = 3; e assim sucessivamente. As tabelas do Anexo A deste documento exibem todos os valores e seus respectivos números de identificação. Para melhor compreensão dos pontos de referência, analise o terceiro gesto (“mau”) juntamente com a Figura 4.9.



(a) Primeira postura e primeiro estado (MAU1) do modelo.



(b) segunda postura e segundo estado (MAU2) do modelo.

Figura 4.9: *Posturas do gesto referente à palavra “mau” em LIBRAS.*

Da mesma forma que foram atribuídos números aos valores das características, serão atribuídas letras às características nas Tabelas 4.3, 4.4 e 4.5. Desta maneira, as letras de A a K correspondem às características PEV-D, PEV-E, PEH-D, PEH-E, CON-D, CON-E, ORI-D, ORI-E, DIP-D, DIP-E e SIB, respectivamente. Abaixo, estão alguns nomes dos gestos, suas siglas entre colchetes, a descrição de como executá-los e pontos de referência que indicam a postura de cada estado do modelo construído.

1) Abafado [ABF]: Mãos verticais abertas, palma a palma, dedos ligeiramente flexionados a cada lado da cabeça (ABF1). Movê-las para diante do rosto e tocá-las pelos dedos mínimos, com expressão de desconforto (ABF2).

Em seguida, com a mão direita, fazer o sinal de A, palma para dentro e apertar o nariz com o polegar e o indicador (ABF3).

	A	B	C	D	E	F	G	H	I	J	K
ABF1	6	6	2	6	20	20	2	2	4	3	2
ABF2	6	6	4	4	20	20	2	2	2	2	2
ABF3	4	9	4	8	1	22	2	3	2	7	2

Tabela 4.3: Configurações modeladas para as características em cada postura do gesto “abafado”.

2) Árvore [ARV]: Braço direito vertical dobrado, mão vertical aberta, palma para frente; mão esquerda aberta, palma para baixo, dedos separados e curvados, cotovelo direito apoiado no dorso da mão esquerda (ARV1). Girar a palma direita para trás, duas vezes (ARV2).

	A	B	C	D	E	F	G	H	I	J	K
ARV1	6	1	3	3	21	21	2	1	1	6	2
ARV2	6	1	3	3	21	21	2	1	2	6	2

Tabela 4.4: Configurações modeladas para as características em cada postura do gesto “árvore”.

3) Mau [MAU]: Mão direita vertical aberta, palma para a esquerda, acima do ombro direito (MAU1). Mover a mão em um arco para a esquerda e para baixo, virando a palma para baixo, com expressão facial negativa (MAU2).

	A	B	C	D	E	F	G	H	I	J	K
MAU1	3	9	2	8	18	22	2	3	4	7	2
MAU2	2	9	5	8	18	22	1	3	6	7	2

Tabela 4.5: Configurações modeladas para as características em cada postura do gesto “mau”.

4.2 Implementações

Foram implementados 8 algoritmos para a realização completa desta trabalho, que serão descritos ao longo desta seção. Eles são exibidos abaixo de maneira introdutória e posteriormente detalhados com maior enfoque.

- Gerador de Observações de Estados: programa que auxilia na etapa de classificação visual das amostras dos gestos, gerando as seqüências de estados observadas;
- Construtor de Modelos: programa responsável pela geração das componentes dos modelos de Markov ocultos;
- Adaptador de Arquivos de Vídeo: programa que converte arquivos de vídeos em formatos diferentes e possibilita a realização de cortes nos mesmos;
- Segmentador de Arquivos de Vídeo: programa que possibilita a segmentação seqüencial de quadros contidos em arquivos de vídeo;
- Extrator de Características de Arquivos de Vídeo: permite a extração de atributos de imagens digitais em arquivos de vídeo;
- Quantizador de Atributos: permite a discretização de atributos com valores reais em conjuntos de amostras;
- Gerador de Observações de Símbolos: programa responsável pela produção de símbolos para os modelos, a partir de características discretas;
- Gerenciador de Modelos: módulo responsável pelas operações de carregamento, armazenamento, consulta e manipulação de HMMs.

Todos os aplicativos implementados foram desenvolvidos com códigos-fonte abertos e gratuitos em linguagem Java, utilizando os pacotes Java Media Framework (JMF), que possibilita a abertura e manipulação de arquivos multimídia, distribuído gratuitamente pela *Sun Microsystems, Inc.*, o ImageJ, uma biblioteca pública e aberta com recursos para análise e processamento de imagens digitais, e o Weka, que se trata de uma coleção de algoritmos implementados de aprendizagem de máquina para tarefas de mineração de dados.

4.2.1 Análise Visual e Obtenção de Componentes Iniciais dos HMMs

Durante a análise visual foram extraídas manualmente as características de cada quadro dos vídeos existentes, com intuito de enquadrar cada quadro do vídeo em um estado reconhecido pelo respectivo modelo ao qual ele pertence. Todos os vídeos referentes aos gestos foram armazenados em um banco de

imagens, para processamento futuro. Por meio do gerador de observações de estados foi possível a navegação pela seqüência de quadros e a associação dos mesmos a um estado do respectivo modelo ao qual o gesto pertence, contido no gerenciador de modelos, e armazena a seqüência de estados em disco, para auxiliar o processo de análise visual.

Com os resultados produzidos pela análise visual, que durou aproximadamente 500 horas, foram alcançadas estatisticamente duas componentes fundamentais de cada modelo: a matriz de probabilidades de transição de estados, denominada A , e o vetor de probabilidades de ocorrência inicial dos estados, representado pelo símbolo π . Cada estado foi representado por uma cadeia de símbolos numéricos que o identificava unicamente dentro do modelo. Com isto, cada arquivo de vídeo produzia uma seqüência de cadeias, sendo que cada cadeia se relaciona diretamente aos quadros existentes no vídeo e a um dos estados do respectivo modelo. Portanto, cada amostra analisada produziu uma seqüência de cadeias, que foram processadas em grupo para obtenção dos valores constituintes de A e π de cada modelo.

As seqüências produzidas pela análise foram tratadas através do construtor de modelos, que, para o momento, gerou as probabilidades de ocorrência inicial e transição de estados para cada um dos gestos, e, em seguida, armazenou os modelos, ainda incompletos, em disco.

4.2.2 Segmentação e Extração de Atributos

Após a etapa de análise, os vídeos foram submetidos aos processos de segmentação e extração de características. Os aplicativos utilizados para esses processos foram o segmentador de imagens e o extrator de características de arquivos de vídeos, em que o primeiro é responsável pela segmentação dos vídeos originais e geração de novos vídeos segmentados, enquanto o segundo é incumbido de realizar a extração das características relevantes dos vídeos segmentados.

Com intuito de realizar a operação de corte nos vídeos, eles foram submetidos ao adaptador de vídeos, que converte as entradas para um formato compatível com o sistema e, posteriormente, habilita a edição do vídeo, produzindo como saída o novo vídeo editado em formato suportado.

Para a segmentação dos quadros foi utilizado o algoritmo baseado em gaussianas, existente na plataforma SIGUS e, posteriormente, foram desenvolvidos outro dois segmentadores, sendo um baseado em combinações de gaussianas e o outro baseado em árvores de decisão. A Figura 4.10 apresenta o resultado comparatório da segmentação utilizando distribuição gaussiana e o resultado obtido com o uso do segmentador utilizando árvore de decisão.

Para auxiliar o processo de segmentação dos vídeos contidos no banco de

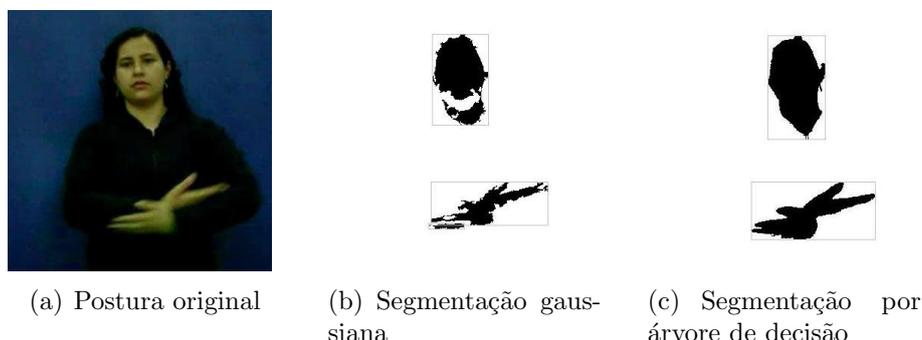


Figura 4.10: *Imagens referentes a uma postura do gesto “baba” antes e após o processo de segmentação. A Figura 4.10(a) é a imagem original; a Figura 4.10(b) é o resultado da segmentação baseada em gaussiana; e a Figura 4.10(c) o resultado da segmentação baseada em árvores de decisão.*

imagens, o segmentador de vídeos possibilita ao usuário a seleção do método segmentador, coleta de amostras para treinamento e visualização experimental da segmentação em quadros específicos do arquivo a ser processado.

Posteriormente, os vídeos, já segmentados, foram enviados à etapa de extração de características. Para realização desta etapa foram extraídos 25 atributos de cada quadro dos vídeos, com auxílio de um analisador de regiões conexas da imagem, contido no ImageJ, e do extrator de vídeos, que automaticamente processou o conjunto completo de gestos contidos no banco de imagens.

De cada imagem foram extraídas a quantidade de partes do corpo presentes, a localização espacial do centro de massa do rosto e os centros de cada mão do usuário normalizados em relação à localização do rosto, os desvios padrões de cada uma dessas partículas, os maiores e menores eixos das elipses correspondentes e seus respectivos ângulos. Esses atributos foram normalizados dentro do aplicativo extrator e armazenados em um arquivo distinto para cada amostra.

4.2.3 Discretização de valores e Geração de Observações

Após a etapa de extração, o sistema conta com um arquivo de características com valores numéricos reais para cada arquivo de vídeo. Através dessas características foram obtidas posteriormente as observações dos modelos de Markov ocultos. No entanto, os modelos de Markov ocultos são de natureza discreta, diferente dos valores obtidos do processo anterior. Por este motivo, é utilizado o quantizador para discretizar os valores reais, a fim de se

obter inteiros correspondentes, sem que haja perda nem sobra excessiva de informação.

O algoritmo utilizado para a quantização implementa um método de conversão de sinais analógicos em digitais. Durante esta etapa foram analisados os valores resultantes do processo de extração com o intuito de alcançar o número adequado de intervalos para cada característica, utilizando como critério as variações existentes em amostras pertencentes ao mesmo gesto, pois tanto a redução quanto o aumento desregrado de intervalos podem interferir de forma negativa no resultado do sistema.

O discretizador carrega todas as características extraídas das amostras e infere os valores discretos para cada amostra existente. Após a discretização, cada conjunto de características, que se relaciona diretamente ao quadro do respectivo vídeo, transforma-se em uma observação para os HMM's através do gerador de observações de símbolos, que é responsável pelo tratamento das características discretas e pela geração das seqüências de observações que são reconhecidas pelos modelos dos gestos.

4.2.4 Obtenção das componentes finais dos HMMs

Dado que um HMM é composto por uma tripla, $\lambda = (\pi, A, B)$, percebe-se que resta apenas a componente B para o preenchimento total dos modelos referentes aos gestos conhecidos pelo sistema. Neste momento, as seqüências de estados provenientes do processo de análise dos vídeos juntamente com as respectivas seqüências de observações, foram processadas por meio do construtor de modelos, e, com isso, a componente B de cada modelo foi alcançada. Dada a condição dos modelos, os requisitos necessários para a experimentação da técnica estão preenchidos.

4.2.5 Classificação das observações

As seqüências de observações alcançadas em um dos processos anteriores foram armazenadas em disco, para que pudessem ser utilizadas nas etapas de classificação e reestimação dos modelos. As observações, no estágio de classificação, eram carregadas e submetidas ao módulo gerenciador de modelos de Markov ocultos, que, entre outras funcionalidades, informa qual o modelo mais provável de ter gerado determinada seqüência de símbolos informada.

Para reconhecer de maneira adequada os gestos existentes na parcela de vídeos reservada para testes, o gerenciador considera a quantidade de símbolos reconhecidos por cada modelo juntamente com as probabilidades de geração de cada um. Desta maneira, seqüências com posturas que por

curtos intervalos de tempo não seguem um comportamento esperado pelo gesto, podem ser reconhecidos de maneira apropriada.

Considerando que o conjunto de símbolos possíveis do sistema possui tamanho aproximado de 210 bilhões, existem diversas configurações possíveis para um mesmo gesto. Dessa forma, o algoritmo distribui em cada modelo, de maneira equiprovável, uma contribuição irrisória aos símbolos que não estavam contidos nas amostras utilizadas para o treinamento dos modelos, alternativa esta que também contribui com a anomalia acima citada. Do mesmo modo, o gerenciador distribui probabilidades mínimas às condições inexistentes às transições e ocorrências iniciais dos estados. Essas probabilidades, por serem naturalmente muito pequenas, não seriam suportadas em variáveis comuns da linguagem escolhida, por isso, foi implementada uma estrutura especial, que apresenta suporte ao armazenamento de números dessa natureza.

O gerenciador de modelos implementado apresenta suporte à improbabilidade acima citada, executando a mesma distribuição equiprovável a cada iteração do algoritmo Baum-Welch, com intuito de possibilitar o aumento de probabilidade dos símbolos irreconhecidos por cada modelo. Porém, os valores dessa distribuição e a ativação de sua execução é parametrizada, possibilitando o procedimento normal, sem adaptações. Entre os valores parametrizados, está o índice de estabilidade entre iterações, que determina o momento de encerramento das iterações de reestimação.

Outra característica peculiar do gerenciador é a alternativa que habilita a mesclagem de amostras de treinamento, que possibilita a execução do método reestimador em várias seqüências de observação em cada iteração, considerando todas as observações antes de maximizar as probabilidades dos modelos.

4.2.6 Diagramas de funcionamento

Construção de modelos

A Figura 4.11 ilustra o funcionamento do processo de construção de modelos do sistema. Os objetos retangulares representam arquivos gravados em disco, cujas extensões/formatos estão indicadas no interior de cada retângulo, enquanto que as elipses são representações dos aplicativos descritos há pouco.

Inicialmente, para questões de compatibilidade entre os vídeos do banco de imagens e as aplicações desenvolvidas, a amostra é enviada em formato MPEG para o adaptador de arquivos, que gera um novo arquivo de saída, em formato MOV. Em seguida, o arquivo produzido pela etapa anterior é enviado aos aplicativos de processamento de imagens e ao gerador de observações de

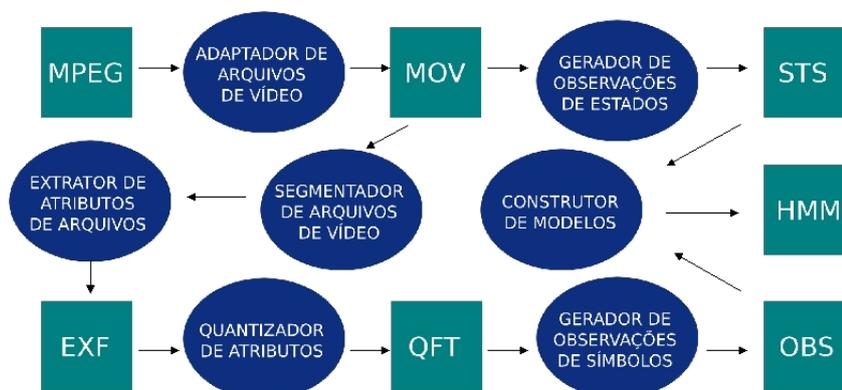


Figura 4.11: Diagrama de representação de funcionamento do processo de construção de modelos.

estados, de maneira simultânea e independente.

Durante o processamento de imagens, as amostras são segmentadas, pelo aplicativo segmentador, e, em seguida, são submetidas ao processo de extração de atributos. Desses processos, é originado um arquivo EXF, que contém as informações extraídas dos quadros da imagem da amostra. Paralelamente, é produzida uma seqüência de observações de estados para cada amostra, que é armazenada em arquivo de formato STS.

As características contidas no arquivo EXF são discretizadas, através do aplicativo quantizador de atributos, e, logo após, rearmazenadas em um arquivo QFT. Por meio desse arquivo de atributos discretos, o gerador de seqüências de símbolos constrói uma seqüência de observação para cada amostra e armazena em um arquivo de extensão OBS.

Dadas as seqüências de estados, STS, e as seqüências de observações, OBS, o aplicativo gerenciador de modelos determina as probabilidades de emissão de símbolos, ocorrência inicial e transição de cada estado e gera o respectivo modelo com essas informações, que é armazenado em formato HMM. Vale observar que para formação de um arquivo HMM são necessárias várias amostras de um mesmo gesto, enquanto que para cada uma das outras extensões é gerado um arquivo por amostra.

Os processos de classificação de amostras e reestimação de parâmetros dos modelos são menos complexos que o processo anterior. Seus diagramas podem ser observados nas Figuras 4.12 e 4.13, respectivamente. Analogamente, os retângulos e elipses têm os mesmos significados que os expostos na explicação do diagrama de construção de modelos. Adicionalmente, o diagrama de classificação de amostras possui o símbolo de uma estrela, e o de reestimação de parâmetros uma circunferência, que representam o modelo

mais provável de gerar a observação de entrada e o novo modelo reestimado para o gesto ao qual a amostra pertence, nesta ordem.



Figura 4.12: *Diagrama de representação de funcionamento do processo de classificação de amostras.*



Figura 4.13: *Diagrama de representação de funcionamento do processo de reestimação de parâmetros dos modelos.*

Capítulo 5

Experimentos

Para a realização das análises e experimentos no sistema, foi criado um banco de imagens referentes aos gestos selecionados [4]. A captura das imagens que compõe este banco foi realizada em um ambiente com fundo estático e uniforme, ou seja, não existiam objetos atrás do usuário e o único objeto com movimento é ele próprio. Foi utilizada uma câmera digital¹ para obtenção das imagens, sendo que esta se manteve fixada em frente ao usuário e imóvel durante as gravações. Os gestos foram executados um por vez e armazenados em arquivos distintos.

Obedecendo a este processo, foram capturadas imagens de 3 colaboradores para os experimentos, em que cada um executou sete ciclos dos 47 gestos selecionados. Com isso, o banco de imagens dedicado à experimentação possuía 21 amostras para cada gesto, totalizando 987 arquivos de vídeo. Durante as gravações não foi utilizado nenhum recurso de apoio para o rastreamento das mãos e da face, como luvas de dados ou câmeras de infra-vermelho.

Para realização dos experimentos, foram divididas as amostras em dois grupos: treinamento e teste. Desta maneira, cerca de 57% do banco foi dedicado ao estágio de treinamento e 43% para a realização dos experimentos. Os experimentos se dividiram em 2 grupos, como descrito abaixo.

O primeiro grupo experimenta a classificação das amostras de treinamento. Ou seja, o sistema apresentaria a classificação dos próprios gestos que foram utilizados para treiná-lo. O segundo grupo verifica a classificação das amostras de teste, tratando as entradas que são desconhecidas pelo sistema. Dentro de cada grupo são realizados dois experimentos: com HMMs constituídos das probabilidades alcançadas matematicamente por fórmula fechada e com HMMs submetidos ao processo de reestimação, utilizando as mesmas amostras de treinamento.

¹Câmera digital, marca Sony, modelo DSC-F828.

Na etapa de reestimação, o sistema executava o algoritmo Baum-Welch, utilizando de maneira simultânea todas as amostras do gesto que foram utilizadas para constituir sua base inicial, com o recurso de mesclagem de observações, disponível no gerenciador de modelos. Esse processo se concluiu no momento em que o índice de variação entre modelo anterior e o modelo reestimado fosse menor que a variação mínima exigida.

As Tabelas 5.2 e 5.3 abaixo representam as matrizes de confusão para cada um dos experimentos realizados com a amostra de treinamento e com a amostra de teste, respectivamente. Nessas tabelas estão apenas os gestos que apresentaram incoerência entre o resultado obtido e o resultado esperado. Ou seja, os gestos que não estão nas tabelas abaixo obtiveram taxa de 100% de acerto. As linhas das tabelas representam o resultado esperado e as colunas o resultado obtido. Para representação dos gestos foram utilizados os mnemônicos descritos na Tabela 5.1.

Após a aplicação do método de reestimação das componentes em todos os modelos reconhecidos pelo sistema, a etapa de classificação retornou exatamente o mesmo resultado de classificação em ambas as amostras. Por este motivo, os resultados foram exibidos em apenas duas tabelas.

Mnemônico	Gesto
ABF	ABAFADO
ABR	ABRIR OS OLHOS
ACR	ACORDAR
ARV	ÁRVORE
BBD	BEBEDOURO
BBL	BÍBLIA
BFR	BUFAR
BGD	BIGODE
BOR	BOIAR
BRR	BURRO
BMV	BEM-VINDO
CDR	CADERNO
CNC	CANCELAR
CRT	CARTAZ
CZN	COZINHAR
DFC	DEFICIÊNCIA AUDITIVA
DFN	DIFUNDIR
ENG	ENGORDAR
ESB	ESBOFETEAR
FND	FUNDAR
FRL	FAROL
GRM	GERMINAR
GTO	GATO
HND	HANDEBOL
HMM	HOMEM
IND	INODORO
INF	INFERIOR
JBT	JABOTICABA
LMB	LAMBUZAR
LBR	LEBRE
MCA	MAÇA
MAU	MAU
MLA	MELÃO

Tabela 5.1: Tabela de relação entre gestos e mnemônicos.

Gesto	Corretos	Incorretos
ACR	10	GTO = 2
BBD	10	DFC = 2
BFR	11	FRL = 1

Tabela 5.2: Resultados obtidos durante classificação da amostra de treinamento.

Gesto	Corretos	Incorretos
ABF	8	BMV = 1
ABR	8	DML = 1
ACR	7	DRR = 2
ARV	6	FRL = 3
BBD	4	CNC = 1 ABF = 1 IND 1 GTO = 1 DFC = 1
BMV	7	FRM = 2
BBL	7	MAU = 1 DFC = 1
BGD	3	DML = 1 LBR = 1 ELS = 3 DFN = 1
BOR	8	FND = 1
BRR	8	CST = 1
CDR	8	CRT = 1
CNC	6	DFC = 3
CRT	8	DRR = 1
CZN	4	GRM = 2 FRM = 3
DFC	8	CNC = 1
DFN	7	ELS = 2
ENG	6	CRT = 1 JNL = 1 DCH = 1
ESB	7	CNC = 1 GTO = 1
FRL	8	MAU = 1
FND	6	GRM = 3
GTO	6	ESB = 2 LMB = 1
GRM	6	CDR = 2 FND = 1
HND	7	HMM = 1 DFC = 1
HMM	2	ABF = 1 MCA = 4 LMB = 1
INF	6	DFC = 2 LMB = 1
IND	7	CNC = 1 DFC = 1
JBT	8	MLA = 1
LMB	8	DFC = 1
LBR	6	ABR = 1 CRT = 1 ELS = 1
MCA	5	CNC = 1 IND = 1 DFC = 1 LMB = 1
MAU	4	ESB = 2 FRM = 1 IND = 1 DFC = 1
MLA	8	HMM = 1

Tabela 5.3: Resultados obtidos durante classificação da amostra de teste.

Capítulo 6

Considerações Finais

A interação homem-máquina guiada por sinais visuais é um rico paradigma de comunicação que amplia o domínio de aplicações computacionais e facilita a utilização dos equipamentos para os seres humanos, principalmente os que possuem limitações na audição e na coordenação motora, tal como os tetraplégicos e surdos. Ferramentas com interação guiada por sinais visuais trazem embutidas consigo o poder de abrigar uma parcela da sociedade que atualmente encontra-se impedida, ou com dificuldades, de utilizar os computadores através das formas habituais de comunicação convencionadas pelo homem, como o mouse e o teclado. Porém, por se tratar de uma tema novo, a utilização deste paradigma ainda não é muito difundida dentro da comunidade de desenvolvedores de aplicações computacionais. Um dos obstáculos para a propagação da aplicabilidade desta interação é a complexidade computacional de implementação destas interfaces, que, em contrapartida, é uma tarefa trivial quando se utiliza meios tradicionais.

Com intuito de vencer o empecilho da complexidade foi desenvolvida uma plataforma de apoio ao desenvolvimento de sistemas para inclusão digital de pessoas com necessidades especiais, o projeto SIGUS. Dentro desta plataforma existem implementações prontas de algoritmos da área de visão computacional, o que auxilia em muito o desenvolvimento deste novo conceito de sistemas computacionais, pois dissipa, quase em totalidade, o óbice anteriormente comentado. Atualmente esta plataforma está bem estruturada e conta com contribuições implementadas das áreas de segmentação, extração de atributos e classificação de imagens, mas ainda possui algumas necessidades. Dentre elas pode-se destacar o reconhecimento de gestos.

Neste trabalho investiu-se na utilização de uma técnica bastante utilizada no reconhecimento de fala e muito bem conceituada, porém, aplicada ao reconhecimento de gestos: os modelos de Markov ocultos (*HMM*). Como dito, a popularização dos modelos de Markov ocultos tem origem no seu

sucesso da sua aplicação no reconhecimento de fala. Nós, utilizamos-no para definir gestos em um sistema reconhecedor, em que cada HMM corresponde a um gesto reconhecido pelo sistema.

O domínio deste sistema reconhecedor é um conjunto de 47 gestos, extraídos do dicionário Trilingüe da Língua Brasileira de Sinais [18, 19]. Estes gestos foram analisados e capturados através de câmeras digitais e armazenadas em um banco de imagens. Foram analisadas as variáveis visuais que melhor caracterizavam os gestos, ou seja, um conjunto de atributos que, em conjunto, descrevem cada gesto de maneira distinta. De fato estas características não descrevem especificamente o gesto, mas sim, cada postura do gesto.

Na análise dos gestos, foram selecionadas as posturas marcantes do gesto, uma vez que um gesto é um seqüência de posturas. Desta maneira, o sistema reconhecedor classificará os gestos utilizando como entrada as observações das posturas existentes na imagem. Após o estabelecimento das posturas chaves dos gestos, estas foram configuradas dentro do conjunto de características acima citado, de maneira empírica, e baseado nestas posturas, foram criados os modelos de Markov ocultos para cada gesto. Entretanto, os modelos construídos possuem apenas os estados, o conjunto de símbolos e as transições, faltando ainda as probabilidades de transição de estados, geração de símbolos e de ocorrência inicial.

Na etapa de obtenção dos vídeos o problema central foi a incompatibilidade entre o equipamento de captura de imagens e o mecanismo de manipulação de vídeos para a linguagem de programação escolhida, causando o desperdício de tempo com o processo de conversão das mídias para um padrão compatível. O segundo problema encontrado nessa etapa foi uma deficiência encontrada na ferramenta de edição de vídeos, o Cinelerra, que não reproduzia todos os quadros do vídeo em imagens, causando falhas notáveis no vídeo resultante.

Durante a etapa de segmentação, pôde ser observado que a análise das colorações de fundo, vestimentas e pele foi uma etapa relevante para o sucesso no resultado do processo, dado que algumas cores que eram visualmente distinguíveis não foram trivialmente discriminadas pelos algoritmos de segmentação utilizados.

Durante a etapa de discretização, o obstáculo maior foi a dispendiosidade em se inferir intervalos apropriados para auxílio no processo de classificação, dado que são diversas características a serem analisadas e diversos quadros. Por mais que na etapa de discretização foi buscada a redução máxima de intervalos, as combinações possíveis de símbolos do sistema era da ordem de 210 bilhões. Desta forma, uma pequena variação em uma das características acarreta na produção de outro símbolo, o que pode prejudicar a classificação

de alguns gestos.

Como as amostras foram divididas em treinamento e testes, é possível que alguns símbolos prováveis em algum gesto não esteja presente na parcela de treinamento. Desta forma, ao ser observado o símbolo ainda não reconhecido, o modelo anula a probabilidade de geração da seqüência, sendo esse um dos empecilhos na etapa de classificação do sistema, que pôde ser resolvido com a distribuição equalitária de uma parcela de probabilidades aos modelos não reconhecidos pelo modelo.

Os experimentos realizados no sistema se limitaram ao domínio de imagens armazenadas em disco. Uma das melhorias que poderia ser adicionada ao trabalho, seria a incorporação de experimentos em imagens capturadas em tempo real, com intuito de abordar dificuldades como variações de iluminação, fundos dinâmicos, etc. Além disso, outra abordagem que enriqueceria o projeto, seriam experimentos com colabores que não estão presentes na parcela de treinamento, para constatar se o sistema suporta independência de usuários.

Outra abordagem interessante seria uma análise comparatória entre o desempenho de classificação do sistema aplicando esta mesma metodologia para a construção dos modelos em relação a modelos gerados aleatoriamente e reestimados através do algoritmo Baum-Welch. Essa análise demonstraria a relevância da identificação das seqüências de estados nas amostras, como foi realizado neste trabalho durante a análise visual.

Enfim, de maneira resumida o sistema apresentou exatidão de 99.11% na classificação das amostras de treinamento e 80.1% nas amostras de teste, sem reestimação de parâmetros dos modelos, e com a aplicação do algoritmo Baum-Welch as taxas de acerto permaneceram as mesmas.

Anexo A

Tabelas de Atributos das Características

PEV	1	entre a cintura e o peito
	2	entre o peito e os ombros
	3	na altura do queixo
	4	na altura da boca
	5	na altura do nariz
	6	na altura dos olhos
	7	acima dos olhos e abaixo do topo da cabeça
	8	acima da cabeça
	9	desconhecida

Tabela A.1: Listagem de valores possíveis de atribuição para a posição espacial vertical da mão (PEV).

PEH	1	à direita, muito afastada do corpo
	2	à direita, afastada do corpo
	3	à direita, próxima ao corpo
	4	ao centro do corpo
	5	à esquerda, próxima ao corpo
	6	à esquerda, afastada do corpo
	7	à esquerda, muito afastada do corpo
	8	desconhecida

Tabela A.2: Listagem de valores possíveis de atribuição para a posição espacial horizontal da mão (PEH).

CON	1	mão em A
	2	mão em S
	3	mão em L
	4	mão em 1
	5	mão em D
	6	mão fechada com o dedo polegar distendido
	7	mão fechada com dedos indicador e polegar distendidos tocando-se
	8	mão fechada com dedos indicador e polegar distendidos em paralelo
	9	mão fechada com todos os dedos distendidos tocando-se
	10	mão em V
	11	mão em U
	12	mão em M
	13	mão em Y
	14	mão em 5
	15	mão em 2
	16	mão em B
	17	mão em C
	18	mão aberta com os dedos unidos
	19	mão aberta com os dedos separados
	20	mão aberta com os dedos unidos e curvados
	21	mão aberta com os dedos separados e curvados
	22	desconhecida

Tabela A.3: Listagem de valores possíveis de atribuição para a configuração da mão (CON).

ORI	1	horizontal
	2	vertical
	3	desconhecida

Tabela A.4: Listagem de valores possíveis de atribuição para a orientação da mão (ORI).

DIP	1	para frente
	2	para trás
	3	para a direita
	4	para a esquerda
	5	para cima
	6	para baixo
	7	desconhecida

Tabela A.5: Listagem de valores possíveis de atribuição para a direção da palma da mão (DIP).

SIB	1	ambas infladas
	2	ambas não infladas

Tabela A.6: Listagem de valores possíveis de atribuição para a stuação das bochechas (SIB).

Referências Bibliográficas

- [1] Y. Ariki and N. Ishikawa. Recognition of head gestures using hidden markov models. In *ICPR '96: Proceedings of the International Conference on Pattern Recognition (ICPR '96) Volume III-Volume 7276*, page 461, Washington, DC, USA, 1996. IEEE Computer Society.
- [2] A. M. Cunha and L. Velho. Hidden markov models. *IEEE Computer Graphics and Applications*, Technical Report TR-02-02, 2002.
- [3] F. Dellaert. The expectation maximization algorithm. Technical Report GIT-GVU-02-20, February 2002.
- [4] J. B. Dias, K. P. Souza, and H. Pistori. Conjunto de treinamento para algoritmos de reconhecimento de libras. *II Workshop de Visão Computacional, São Carlos*, October 2006.
- [5] S. Eickeler, A. Kosmala, and G. Rigoll. Hidden Markov Model Based Continuous Online Gesture Recognition. In *Int. Conference on Pattern Recognition (ICPR)*, pages 1206–1208, Brisbane, 1998.
- [6] A. El-Yacoubi, R. Sabourin, M. Gilloux, and C. Y. Suen. Off-line handwritten word recognition using hidden markov models. pages 191–230, 1999.
- [7] W. Feller. *Introdução à teoria das probabilidades e suas aplicações*. Editora Edgard Blücher Ltda., 1976.
- [8] C. Lee and Y. Xu. Online, interactive learning of gestures for human/robot interfaces. *1996 IEEE International Conference on Robotics and Automation*, 4:2982–2987, 1996.
- [9] S. Lipschutz. *Probabilidade*. Makron Books, 1993.
- [10] M. Morita and L. S. Oliveira. Introdução aos modelos escondidos de markov. Technical report, PPGIA-PUCPR, Curitiba-Brazil, November 1998.

- [11] B. A. Myers. Why are Human-Computer Interfaces Difficult to Design and Implement? Technical Report CMU-CS-93-183, July 1993.
- [12] Y. Nam and K. Wohn. Recognition of space-time handgestures using hidden markov model. *ACM Symposium on Virtual Reality Software and Technology*, pages 51–58, 1996.
- [13] S. C. W. Ong and S. Ranganath. Automatic sign language analysis: A survey and the future beyond lexical meaning. *IEEE Trans. Pattern Anal. Mach. Intell.*, 27(6):873–891, 2005.
- [14] H. Pistori. *Tecnologia Adaptativa em Engenharia de Computação: Estado da Arte e Aplicações*. PhD thesis, Universidade de São Paulo, São Paulo, São Paulo, Brasil, 2003.
- [15] H. Pistori, P. S. Martins, M. C. Pereira, and J. J. Neto. Sigus - plataforma de apoio ao desenvolvimento de sistemas para inclusão digital de pessoas com necessidades especiais. *IV Congresso Iberdiscap: Tecnologias de Apoio a Portadores de Deficiência*, Vitória, February 2006.
- [16] D. Prescher. A tutorial on the expectation-maximization algorithm including maximum-likelihood estimation and em training of probabilistic context-free grammars. *ArXiv Computer Science e-prints*, December 2004.
- [17] L. R. Rabiner. A tutorial on hidden markov models and selected applications in speech recognition. *IEEE Computer Graphics and Applications*, 77(2):42–53, 1989.
- [18] F. C. Capovilla; W. D. Raphael. *Dicionário Enciclopédico Ilustrado Trilíngüe da Língua de Sinais Brasileira*, volume Volume I: Sinais de A a L. São Paulo, SP: Edusp, Imprensa Oficial, 2002.
- [19] F. C. Capovilla; W. D. Raphael. *Dicionário Enciclopédico Ilustrado Trilíngüe da Língua de Sinais Brasileira*, volume Volume I: Sinais de M a Z. São Paulo, SP: Edusp, Imprensa Oficial, 2002.
- [20] W. Sandler and D. Lillo-Martin. *Natural Sign Languages*, pages 533–562. Oxford: Blackwell, 2001.
- [21] T. Starner. Visual recognition of american sign language using hidden markov models. Technical Report Master’s Thesis, MIT, Feb 1995, Program in Media Arts & Sciences, MIT Media Laboratory, 1994.

-
- [22] D. Tanguay. Hidden markov models for gesture recognition. Technical report, Cambridge: Massachusetts Institute of Technology, Department of Electrical Engineering and Computer Science., 1993.
- [23] S. Theodoridis and K. Koutroumbas. *Pattern recognition*. Academic press, 1999.
- [24] A. D. Wilson and A. F. Bobick. Hidden markov models for modeling and recognizing gesture under variation. *IJPRAI*, 15(1):123–160, 2001.
- [25] C. A. Ynoguti. *Reconhecimento de Fala Contínua Usando Modelos Ocultos de Markov*. PhD thesis, Faculdade de Engenharia Elétrica e de Computação, Universidade Estadual de Campinas, Campinas, São Paulo, Brasil, 1999.