Alexandre Savaris

Avaliação comparativa de técnicas para reconhecimento de gestos estáticos e dinâmicos com foco em precisão e desempenho

Florianópolis – SC
2010

UNIVERSIDADE FEDERAL DE SANTA CATARINA
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO

Alexandre Savaris

Avaliação comparativa de técnicas para reconhecimento de gestos estáticos e dinâmicos com foco em precisão e desempenho

Dissertação submetida à Universidade Federal de Santa Catarina como parte dos requisitos para a obtenção do grau de Mestre em Ciência da Computação.

Prof. Dr. rer.nat. Aldo von Wangenheim

Florianópolis, fevereiro de 2010

Catalogação na fonte pela Biblioteca Universitária
da
Universidade Federal de Santa Catarina

S256a Savaris, Alexandre
Avaliação comparativa de técnicas para reconhecimento de gestos estáticos e dinâmicos com foco em precisão e desempenho [dissertação] / Alexandre Savaris ; orientador, Aldo von Wangenheim. - Florianópolis, SC 2010.
100 p.: il., grafs., tabs.

Dissertação (mestrado) - Universidade Federal de Santa Catarina. Centro Tecnológico. Programa de Pós-Graduação em Ciência da Computação.

Inclui referências

1. Ciência da computação. 2. Gestos. 3. Reconhecimento. 4. Postura humana. 5. Trajetória. 6. Interação homem-máquina. I. Wangenheim, Aldo v. (Aldo von). II. Universidade Federal de Santa Catarina. Programa de Pós-Graduação em Ciência da Computação. III. Título.

CDU 681

Avaliação comparativa de técnicas para reconhecimento de gestos estáticos e dinâmicos com foco em precisão e desempenho

Alexandre Savaris

Esta dissertação foi julgada adequada para a obtenção do título de Mestre em Ciência da Computação, área de concentração de Sistemas de Computação, e aprovada em sua forma final pelo Programa de Pós-Graduação em Ciência da Computação.

Coordenador

Dr. Mário Antonio Ribeiro Dantas
Universidade Federal de Santa Catarina

Orientador

Dr. rer.nat. Aldo von Wangenheim
Universidade Federal de Santa Catarina

Banca Examinadora

Dr. Renato Fileto
Universidade Federal de Santa Catarina

Dr. rer.nat. Eros Comunello
Universidade Federal de Santa Catarina

Dra. Luciana Porcher Nedel
Universidade Federal do Rio Grande do Sul

Dr. Luiz Felipe de Souza Nobre
Universidade Federal de Santa Catarina

“Time discovers truth.”
Seneca

À toda a minha família, pelo apoio.
À Milena, pelo companheirismo.
Ao Eros, pelo convite que resultou em meu ingresso no grupo Cyclops.
Ao prof. Aldo, pela confiança refletida nas responsabilidades crescentes junto ao grupo.
Aos colegas do LAPIX e LABTELEMED.

Sumário

Lista de figuras

Lista de quadros

Lista de tabelas

Lista de acrônimos e símbolos

2D	Bidimensional
3D	Tridimensional
AFD	Autômato Finito Determinístico
AGR	Accelerometer Gesture Recognizer
ANN	Approximate Nearest Neighbor
API	Application Programming Interface
CAD	Computer-Aided Design
CTRNN	Continuous Time Recurrent Neural Network
DOF	Degrees of Freedom
DP	Desvio padrão
EP1	Estratégia de particionamento 1
EP2	Estratégia de particionamento 2
FANN	Fast Artificial Neural Network
GPU	Graphics Processing Unit
HCI	Human-Computer Interaction
HMHH	Hierarchical Motion History Histogram
HMM	Hidden Markov Model
IHC	Interação Humano-Computador
JAST	Joint Action Science and Technology
LRB	Left-Right Banded
LSH	Locality Sensitive Hashing
MEB	Minimum Enclosing Ball
MHI	Motion History Image
MM	Markov Model
RBF	Radial Basis Function
SGONG	Self-Growing and Self-Organized Neural Gas

SVM	Support Vector Machine
TMA	Tempo médio de avaliação
TMT	Tempo médio de treinamento
UFSC	Universidade Federal de Santa Catarina
δ	Delta
Θ	Teta
Σ	Sigma
Φ	Fi

Resumo

É um comportamento comum aos seres humanos utilizar gestos como forma de expressão, como um complemento à fala ou como uma forma auto-contida de comunicação. No campo da Interação Humano-Computador, esse comportamento pode ser adotado na construção de interfaces alternativas, objetivando facilitar o relacionamento entre os elementos humano e computacional. Atualmente, várias técnicas para reconhecimento de gestos são descritas na literatura; porém, as validações dessas técnicas são executadas de maneira isolada, o que dificulta a comparação entre elas. Para reduzir essa lacuna, este trabalho apresenta uma comparação entre técnicas estabelecidas para o reconhecimento de gestos estáticos (posturas) e gestos dinâmicos (trajetórias). Essas técnicas são organizadas de forma a avaliar um conjunto de dados comum, adquirido por meio de uma luva instrumentada e um rastreador de movimento, gerando resultados em termos de precisão e desempenho. Especificamente para trajetórias, o processo de avaliação considera técnicas conhecidas (redes neurais e modelos ocultos de Markov) e uma nova heurística baseada em autômatos finitos determinísticos, idealizada e desenvolvida pelos autores. Os resultados obtidos mostram que o classificador baseado em uma SVM (*Support Vector Machine*) apresentou a melhor generalização, com as melhores taxas de reconhecimento para posturas. Para trajetórias, por sua vez, o classificador baseado em uma rede neural gerou os melhores resultados. Em termos de desempenho, todos os métodos apresentaram resultados suficientemente rápidos para serem usados de forma interativa. Finalmente, o presente trabalho identifica e discute um conjunto de critérios relevantes que deve ser observado nas etapas de construção, treinamento e avaliação dos classificadores, e sua relação com os resultados finais.

Abstract

It is a common behavior for human beings to use gestures as a means of expression, as a complement to speaking, or as a self-contained communication mode. In the field of Human-Computer Interaction, this behavior can be adopted to build alternative interfaces, aiming to ease the relationship between the human element and the computational element. Currently, various gesture recognition techniques are described in the literature; however, the validation studies of these techniques are usually performed isolatedly, which difficult comparisons between them. To reduce this gap, this work presents a comparison between well-established techniques used in the recognition of static gestures (postures) and dynamic gestures (trajectories). These techniques evaluate a common dataset, acquired from an instrumented glove and a motion tracker, and generate results for precision and performance measurements. Specifically for trajectories, the evaluation process considers known techniques (neural networks and hidden Markov Models) and a new heuristic based on deterministic finite automata, designed and developed by the authors. The results obtained show that the classifier implemented as a Support Vector Machine (SVM) presented the best generalization, with the highest recognition rate for postures. For trajectories, in turn, a neural network achieved the best results. In terms of performance, all methods presented evaluation times fast enough to be used interactively. Finally, this work identifies and discusses a set of relevant criteria that must be observed in the stages of construction, training and evaluation of the classifiers, and its relation to the final results.

1. Introdução

Interfaces baseadas em gestos oferecem alternativas às formas tradicionais de interação entre seres humanos e computadores, largamente apoiadas no par teclado/*mouse*. Para que essas interfaces sejam construídas e disponibilizadas, uma série de quesitos deve ser atendida; dentre eles, pode-se destacar o processo de reconhecimento de gestos – responsável por coletar dados, reconhecer e classificar esses dados como gestos válidos e mapeá-los a eventos ou comandos de aplicação. Este capítulo contextualiza o presente trabalho em relação ao processo supracitado, explicitando seus objetivos e sua organização.

1.1 Contextualização

A disciplina de Interação Humano-Computador (IHC), ou *Human-Computer Interaction* (HCI) trata do projeto, implementação e avaliação de alternativas para o interfaceamento entre o elemento humano e o elemento computacional (ACM SIGCHI, 2009). É tida como uma área de estudo multidisciplinar, envolvendo ciência da computação, psicologia, sociologia, antropologia e *design* industrial, dentre possíveis outros ramos de conhecimento. Cada ramo de conhecimento envolvido assume um ponto de vista específico, de acordo com seu histórico e atuação. Apesar de independentes entre si, esses pontos de vista são tomados em conjunto para prover subsídios objetivando o estabelecimento de técnicas de interação. Essas técnicas voltam-se ao relacionamento entre seres humanos – no sentido individual ou coletivo – e recursos computacionais, identificados por um sem-número de dispositivos de *hardware* e *softwares* aplicativos.

Nas últimas décadas o estudo, o desenvolvimento e a aplicação de técnicas de interação estabeleceram marcos que nortearam a forma como se dá a relação entre homem e máquina atualmente. Dentre esses marcos, podem ser destacados como principais a criação de técnicas para manipulação direta de objetos gráficos, o desenvolvimento do *mouse* como dispositivo de apontamento, a criação de interfaces baseadas em janelas, o aperfeiçoamento de aplicativos de desenho, edição de texto, planilhas de cálculo e projeto assistido por computador – *Computer-Aided Design* (CAD), a disseminação do hipertexto e a evolução dos vídeo-games (MYERS, 1998). Como complemento aos marcos citados, podem ser relacionadas técnicas de interação baseadas em multimídia, representações tridimensionais, realidade virtual, reconhe-

cimento de linguagem natural e, não menos importante, reconhecimento de gestos.

Gestos podem ser definidos como movimentos executados pelo corpo ou partes do corpo de uma pessoa (como braços e pernas, por exemplo), objetivando expressar ou enfatizar uma idéia, sentimento ou atitude (MERRIAM-WEBSTER, 2009). Como forma de expressão, os gestos podem ser utilizados em complemento à comunicação verbal, ou como uma forma de comunicação autônoma – bem identificada através das linguagens de sinais. No contexto da disciplina de IHC, os gestos fornecem uma forma diferencial de interação através da qual um ambiente computacional controlado pode ser operado; essa operacionalização se dá pelo mapeamento de gestos para funções de aplicação, de forma a aproveitar a naturalidade dos mesmos para o controle do ambiente (PAVLOVIĆ et al., 1997). Apesar de existir uma definição genérica para a disciplina, há uma dependência relativa à complexidade das aplicações e dos correspondentes processos interativos para a caracterização de gestos. Essa dependência permite que se adotem especializações à definição genérica, adaptando-a a cada caso.

A construção de uma interface gestual compreende uma série de etapas, dentre as quais se destaca o processo de reconhecimento e classificação de gestos. É nessa etapa que os gestos executados pelos usuários são interceptados, avaliados e interpretados, sendo traduzidos para comandos que serão repassados à aplicação, objetivando controlá-la. A necessidade de reconhecimento e classificação leva a uma primeira questão: como será possível executar a interpretação corretamente, de forma a evitar o reconhecimento de gestos indevidos e, ainda, garantir que todos os gestos relevantes sejam considerados? Atualmente, diversas técnicas e algoritmos podem executar essa tarefa. Há, porém, questões secundárias – mas não menos importantes – que surgem no momento da escolha da técnica a ser utilizada. Qualquer técnica pode ser utilizada para qualquer vocabulário de gestos? Uma técnica pode ser considerada genérica o suficiente para ser usada em todas as situações? Uma técnica específica é capaz de gerar resultados em tempo hábil, de forma a permitir a utilização da interface relacionada em tempo real? A técnica escolhida se adapta a qualquer tamanho de vocabulário, e é independente das características que compõem esse vocabulário? Respostas a essas questões não são encontradas explicitamente em trabalhos relacionados a reconhecimento de gestos, o que dificulta a escolha de técnicas adequadas e gera dúvidas quanto à aplicabilidade de interações gestuais.

1.2 Objetivo

Conforme explicitado na contextualização, há um conjunto de questões que circundam a escolha de um método para o reconhecimento de gestos. A falta de respostas diretas a essas questões dificulta a construção de interfaces gestuais, dadas as numerosas opções existentes em termos de técnicas e algoritmos.

Visando fornecer respostas às perguntas citadas anteriormente, o objetivo do presente trabalho é comparar um conjunto de técnicas para reconhecimento de gestos em termos de precisão e desempenho. As técnicas selecionadas para comparação foram escolhidas pela sua relevância, tendo sido extensivamente estudadas e utilizadas em experimentos relacionados a interfaces alternativas. Especificamente para o reconhecimento de trajetórias, a comparação foi feita entre técnicas descritas na literatura e uma heurística definida pelos autores, fundamentada no modelo de autômatos finitos determinísticos. Como principal contribuição, este trabalho identifica o melhor conjunto de métodos para o reconhecimento de posturas e trajetórias derivadas de dados instrumentados, restritas a um vocabulário previamente conhecido, bem como a melhor parametrização desses métodos, definida empiricamente.

Para que os métodos pudessem ser comparados, foi necessária a construção de um vocabulário de gestos, baseado em trabalhos relacionados e restrito pelas limitações do *hardware* utilizado. O trabalho descreve o processo de construção do vocabulário, desde a sua concepção até o método utilizado para a aquisição dos dados que o compõem; é descrita, também, a etapa de comparação entre os métodos escolhidos, com a tabulação dos dados de interesse e as devidas considerações sobre os resultados obtidos.

1.3 Organização do trabalho

A presente dissertação está organizada em capítulos, como segue. Definições sobre tipos de gestos e estruturas de classificação são apresentadas no capítulo dois. No capítulo três são apresentados detalhes sobre o processo de reconhecimento de gestos, e no capítulo quatro são descritas técnicas de *benchmarking* para interfaces baseadas em gestos. No capítulo cinco são listados os trabalhos relacionados, base para a escolha dos métodos de reconhecimento utilizados na comparação. O capítulo seis descreve o ambiente experimental, que compreende a modelagem e a aquisição dos dados do vocabulário utilizado no trabalho

e o *hardware*/*software* utilizados no decorrer do processo. Os capítulos sete e oito apresentam, respectivamente, os resultados obtidos e uma discussão acerca desses resultados, e o capítulo nove relata as conclusões finais e relaciona um conjunto de possíveis trabalhos futuros.

2. Gestos como forma de interação

A adoção de gestos como meio de interação com computadores objetiva tornar o contato entre homem e máquina mais natural. Interfaces mais intuitivas e menos intimidadoras contribuem para que os usuários possam usufruir das facilidades oferecidas pelos *hardwares* e *softwares* disponíveis. A construção de uma interface com essas características implica em uma série de decisões, que vão desde o rascunho inicial que idealiza o modo de funcionamento da mesma, até os testes a serem executados que a validem. Neste capítulo, serão apresentados conceitos, classificações, descrições de processo e estratégias de *benchmarking* úteis no entendimento da construção de uma interface gestual.

2.1 Caracterização dos termos postura e gesto

No contexto do desenvolvimento de interfaces, o termo *gesto* é comumente utilizado como uma generalização para uma forma diferencial de interação. Quanto à sua composição, é possível identificar elementos específicos passíveis de serem avaliados de forma independente ou em conjunto. Segundo LAVIOLA (1999), uma classificação quanto ao dinamismo do gesto pode ser estabelecida, resultando na seguinte diferenciação:

- *posturas* (também conhecidas como gestos estáticos) são definidas como posicionamentos de partes do corpo de uma pessoa, relacionados com um momento em uma linha de tempo; podem ser divididas em posturas simples e posturas complexas;
- *gestos* (também conhecidos como gestos dinâmicos) são definidos como movimentos relacionados a trajetórias, relacionados ou não com um intervalo em uma linha de tempo; podem ser divididos em gestos simples e gestos complexos[1]. A diferença entre posturas e gestos pode ser visualizada na Fig. 1.

As possíveis combinações entre posturas e gestos são amplamente exploradas e contextualizadas de acordo com os trabalhos desenvolvidos. Assim, diferentes objetivos são atingidos através da adaptação de posturas e gestos à realidade de cada aplicação. IWAI et al. (1999) utiliza os gestos simples dos braços de uma pessoa como meio de interação, validando sua proposta através do reconhecimento de

[1] Na literatura, gestos dinâmicos também são definidos como sequências de posturas executadas em um intervalo de tempo. Neste trabalho, é adotada a definição dada por LaViola (1999).

mímicas relacionadas a instrumentos musicais e saudações definidas na linguagem japonesa de sinais. Nesse contexto, as posturas são utilizadas como delimitações para o início e o fim do gesto, não possuindo um significado quando avaliadas em separado. O trabalho de LEE et al. (1998), por sua vez, utiliza tanto posturas quanto gestos para controlar *avatares* em um ambiente virtual. Cada postura significativa é relacionada a uma ação, que permite a movimentação ou a interação dos *avatares* entre si. Os gestos são utilizados como um complemento às posturas, indicando direções e orientações no espaço virtual. Já o trabalho de TANI et al. (2007) atribui significância tanto para posturas quanto para gestos, em uma aplicação utilizada na visualização e manipulação de imagens radiológicas. Nesse exemplo, os eventos da aplicação podem ser mapeados para uma determinada postura, um determinado gesto ou um par formado por uma postura e um gesto.

Figura 1: Exemplos de posturas (à esquerda) e gesto (à direita) (DIPIETRO et al., 2008).

Os trabalhos citados permitem visualizar diferentes possibilidades para a utilização de gestos em contextos específicos. Em cada trabalho, coube aos autores a atribuição de significados a posturas, gestos ou ao conjunto formado por posturas e gestos. Essa situação é facilmente identificável em trabalhos correlatos, onde não se distingue um padrão estipulado a ser seguido no tocante à relação postura/significado ou gesto/significado. A liberdade de escolha de posturas e gestos para cada aplicação, individualmente, é salutar e pode permitir que cada *software* seja operado de forma específica, possibilitando inclusive que cada usuário configure uma forma particular de interação.

Há, porém, situações nas quais a adoção de uma classificação é interessante. A identificação de semelhanças entre posturas e gestos, com seu posterior agrupamento, permite que sejam definidas estratégias de reconhecimento e tratamento que extrapolam o individual e que po-

dem ser aplicadas ao coletivo. A classificação de tais grupos é feita através de taxonomias, sendo conhecidas diferentes propostas que objetivam sua estruturação e organização.

2.2 Taxonomias para a organização e classificação de posturas e gestos

O conjunto de gestos – ou vocabulário – a ser utilizado em sessões interativas pode variar de aplicação para aplicação. Dessa forma, é possível atribuir uma identidade única a cada interface, sendo que essa unicidade é garantida pela escolha de gestos específicos relacionados a eventos ou ações. Apesar da liberdade garantida por essa definição, o agrupamento de gestos tidos como semelhantes (sob algum critério) é uma boa prática que permite o entendimento e a aplicação de técnicas específicas de tratamento para os mesmos. Comumente, esse agrupamento é feito através de taxonomias. Uma taxonomia para gestos corresponde a uma classificação feita sob um conjunto de critérios, que permite identificar semelhanças e estabelecer relações baseadas em estrutura ou significado. Existem diferentes propostas de taxonomias na literatura, cada qual com a sua abrangência e foco, porém com um objetivo comum: permitir que gestos semelhantes sejam classificados como tal e, posteriormente, tratados de forma análoga.

As primeiras taxonomias definidas para gestos datam de trabalhos da década de quarenta, e não apresentam relação com tecnologias ou técnicas computacionais. Esses trabalhos foram desenvolvidos por linguistas, neurologistas e terapeutas com a intenção de reproduzir a fala de acordo com as funções cerebrais e os processos cognitivos dos indivíduos; tais estudos foram (e são) considerados como voltados ao desenvolvimento das capacidades de comunicação individual dos envolvidos.

WEXELBLAT (1994), em sua dissertação de mestrado, descreve e apresenta um conjunto de quatro taxonomias precursoras na área da classificação de gestos. As taxonomias são organizadas em quatro ou cinco eixos de classificação, e são baseadas fundamentalmente na observação de interações interpessoais e em experiências realizadas em ambientes controlados. Tanto as observações quanto as experiências realizadas utilizam os gestos como uma forma paralela de expressão (e não como uma forma complementar) quando associados à fala. O Quadro 1 permite visualizar os eixos propostos pelas quatro taxonomias, bem como as relações existentes entre as diferentes propostas.

Quadro 1: Propostas de taxonomias para gestos.

	Propostas				Características
	Kendon	**McNeill & Levy**	**Rime & Schiaratura**	**Efron**	
Eixos	Fisiográfico	Icônico	Fisiográfico	Cinetográfico	Representação pictórica (1)
	Ideográfico	Metafórico	Icônico	Ideográfico	Representação de idéias (2)
	Gesticulação	Batida / *Buterworths*	Marcação	Batida	Ritmo do diálogo (3)
	Gesto autônomo	Simbólico	Simbólico	Simbólico / Emblemático	Significado auto-contido (4)
	-	Dêitico	Dêitico	-	Indicação e apontamento (5)

Fonte: (WEXELBLAT, 1994) – adaptação.

As características que norteiam a classificação dos gestos nos quatro ou cinco eixos presentes no Quadro 1 podem ser detalhadas como segue:

- (1) permite criar uma *representação figurada* dos objetos sobre os quais trata o diálogo; os objetos representados são componentes obrigatórios do conteúdo do diálogo;
- (2) retrata as idéias expostas pelo interlocutor, sem que as mesmas necessitem ser representadas através de objetos bem definidos;
- (3) marca o ritmo do diálogo, através da enfatização de partes da conversa, introdução de novos elementos ou divisão do conteúdo abordado em tópicos;
- (4) apresenta um significado direto, que independe da forma adotada. Exemplos: dedos indicador e médio formando o sinal de vitória; dedos polegar e indicador, em círculo, representando o sinal de *OK*. Não necessita de um complemento verbal para ser compreendido;
- (5) usado para a indicação e apontamento de uma pessoa ou área de interesse; leva em conta o espaço que cerca o interlocutor.

Com a vinculação dos gestos às interfaces multimodais, surgiram diferentes taxonomias que abrangeram também as áreas tecnológicas e de domínio de aplicação. Uma proposta bastante abrangente é dada por KARAM et al. (2005), na qual a taxonomia definida é resultado de uma revisão de literatura de trabalhos publicados em um intervalo de 40 anos sobre processos de interação baseados em gestos. A principal conclusão

dos autores é de que gestos existem em diferentes formas para diferentes domínios de aplicação, e que é o domínio de aplicação o responsável por determinar os dispositivos de entrada e saída a serem utilizados. A taxonomia, que pode ser visualizada na Fig. 2, é dividida em quatro eixos principais: *domínios de aplicação* (onde, efetivamente, as interfaces de gestos são utilizadas), *tecnologias de entrada e saída* (o que se utiliza para adquirir as informações que compõem os gestos, e como essas informações podem ser percebidas e visualizadas), *respostas dos sistemas* (como o *feedback* é dado aos usuários) e *estilos de gestos* (quais características são perceptíveis e utilizáveis como base de classificação para diferentes gestos executados). Dentre os eixos citados, o responsável pelo agrupamento dos estilos de gestos merece um maior destaque por tratar especificamente dos gestos em si, sem relações diretas com tecnologias específicas. Sua divisão pode ser descrita como segue.

- *Gestos dêiticos* – gestos que permitem estabelecer identidades ou localizações espaciais de objetos no contexto do domínio de aplicação. No domínio de aplicação *desktop*, por exemplo, podem ser utilizados como forma de escolha de objetos virtuais passíveis de manipulação.
- *Gesticulação* – gestos executados, normalmente, como um suporte à comunicação verbal. Não possuem um padrão definido, sendo altamente dependentes do contexto no qual estão inseridos. São componentes de interfaces multimodais onde o domínio de aplicação, a fala e a execução de gestos se complementam para permitir a interação. Exemplos incluem gestos que buscam enfatizar uma opinião dada sobre um assunto em discussão.
- *Manipulação* – gestos que permitem controlar um objeto ou entidade, estabelecendo um relacionamento entre a mão/braço do executor do gesto e o objeto/entidade. Podem ser classificados quanto aos graus de liberdade (DOF – *degrees of freedom*) possíveis (por exemplo, *displays* bidimensionais – 2D), número de dimensões (2D e 3D, extrapolando inclusive as dimensões espaciais e assumindo outras, como temperatura ou resistência), a combinação de ambos e o mapeamento entre objetos físicos e virtuais, fazendo com que alterações aplicadas aos primeiros reflitam nos últimos. A taxonomia define que, para que uma manipulação seja interpretada como um gesto, um evento ou ação deve estar vinculado a essa manipulação.

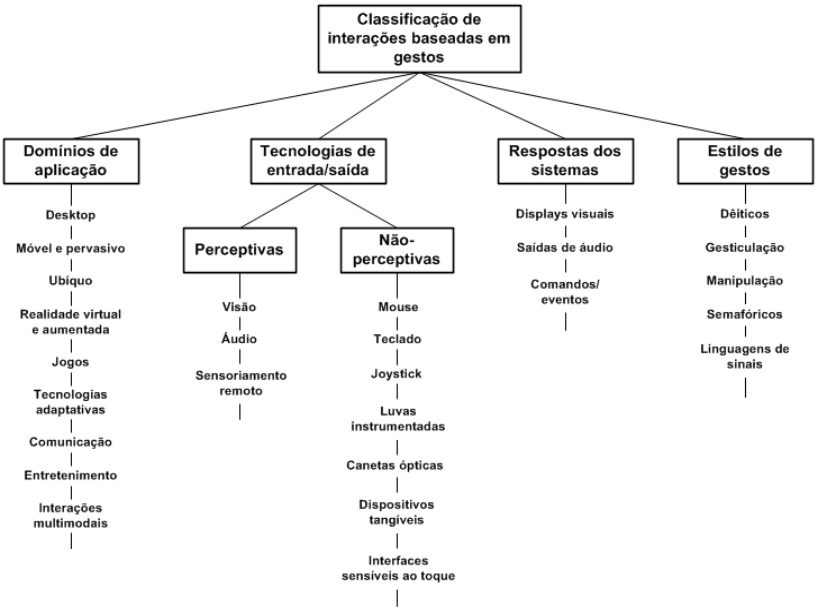

Figura 2: Proposta de taxonomia para classificação de gestos (KARAM et al., 2005) – adaptação.

- *Gestos semafóricos* – gestos pertencentes a um vocabulário pré-definido, vinculados a eventos ou ações. São amplamente utilizados como uma forma de comunicação. Exemplos: gesto representando *OK* (dedos polegar e indicador formando um círculo) e adeus (movimento da mão de um lado para o outro).
- *Linguagens de sinais* – gestos que se diferenciam dos gestos semafóricos e da gesticulação em si por apresentarem uma estrutura léxica e gramatical bem definida, e por serem orientados à comunicação e não à execução de eventos ou ações. Exemplo: *libras*.

Propostas mais simples, normalmente focadas em um domínio de aplicação específico, também podem ser citadas. A taxonomia definida por NEHANIV et al. (2005a e 2005b), por exemplo, propõe uma classificação de gestos no domínio de aplicação da robótica, objetivando inferir as intenções dos gestos executados. Os trabalhos assumem que as técnicas bem conhecidas da área de IHC podem ser utilizadas para o reconhecimento mecânico dos gestos; porém, a contextualização da execução dos mesmos ainda é tema de estudos e fator imprescindível

para que a intenção de um gesto possa ser identificada. Na taxonomia, são definidas cinco classes de gestos:

- *gestos irrelevantes ou manipulativos* – executados com o objetivo de manipular objetos presentes no ambiente ou decorrentes de movimentos corporais naturais. Não representam uma tentativa de comunicação ou interação. Exemplos: mover os braços durante uma caminhada, pegar um copo para beber, entre outros;
- *expressão de comportamento* – objetivam auxiliar na comunicação com outras pessoas, não respeitando regras ou convenções. Exemplos: gestos enfáticos vinculados à defesa de uma idéia ou opinião durante uma discussão;
- *gestos simbólicos* – buscam contribuir para uma comunicação através de símbolos bem-definidos e previamente conhecidos pelos participantes do processo comunicativo. Exemplos: acenos e saudações;
- *interação* – gestos usados especificamente para regular a interação com parceiros em um ambiente colaborativo. Possibilitam iniciar, manter, sincronizar, organizar e terminar a interação, tendo como base a existência de um emissor e de um receptor. Exemplos: estender a mão para solicitar um objeto ou estender a mão para entregar/oferecer um objeto;
- *referenciais e apontamento* – gestos utilizados para identificação (tanto de objetos quanto de indivíduos) no ambiente.

Os trabalhos ressaltam a possibilidade de um mesmo gesto ser classificado em mais de um grupo, o que leva à necessidade de inferir sobre o contexto no qual o mesmo foi executado para identificar seu real significado. Além disso, as possíveis ambiguidades e diferenças são consideradas (como, por exemplo, o mesmo gesto ser executado com objetivos completamente distintos, e o mesmo gesto possuir interpretações diversas de acordo com a cultura dos envolvidos, em escala interracial ou regional).

Ainda tendo os domínios de aplicação como base para a estruturação e construção de taxonomias, o trabalho de WOBBROCK et al. (2009) propõe uma classificação taxonômica para gestos voltados a superfícies sensíveis ao toque. O trabalho inova por transferir a responsabilidade pela criação dos gestos aos usuários; para isso, são apresentados os resultados da execução de gestos hipotéticos, e é solicitado aos usuários que executem o gesto que lhes parecer mais conveniente para que se atinja o resultado apresentado. A taxonomia

divide-se em quatro eixos principais: *forma* (postura assumida pelas mãos e o número de pontos de contato utilizados), *natureza* (diferenciando gestos simbólicos e gestos manipulativos), *ligação* (definindo o relacionamento entre os objetos manipulados e a representação de mundo no qual estão inseridos) e *fluxo* (vinculando as respostas do sistema aos gestos no término de sua execução ou no período durante o qual o gesto é executado). Cada um dos eixos é subdividido em categorias menores, sendo possível ainda subdividir as categorias através da execução dos gestos utilizando uma ou duas mãos.

A avaliação das propostas apresentadas permite identificar pontos de intersecção, onde os mesmos tipos de gestos são classificados analogamente, porém com uma nomenclatura diversa. Os gestos semafóricos definidos por KARAM et al. (2005), por exemplo, correspondem aos gestos simbólicos encontrados nos trabalhos de NEHANIV et al. (2005a e 2005b). Essas intersecções, além de permitirem uma comparação entre diferentes propostas, possibilitam também que modelos híbridos sejam construídos, centrados em eixos comuns e complementados por eixos secundários voltados especificamente a determinados domínios de aplicação.

Os trabalhos supracitados demonstram que o processo de organização de gestos pode levar em conta aspectos humanos, aspectos tecnológicos e domínios de aplicação, individual ou coletivamente. É importante salientar que, independentemente desses aspectos, os objetivos continuam convergindo para o agrupamento de gestos tidos como semelhantes. Tal agrupamento é interessante para o processo de reconhecimento dos gestos executados, visto que a escolha de técnicas para captura, tratamento e interpretação pode ser feita com base nas características comuns pertinentes a grupos de gestos específicos. O capítulo a seguir descreve esse processo em detalhes.

3. O processo de reconhecimento de gestos

A etapa de modelagem de uma interface baseada em gestos corresponde à definição do domínio de aplicação, juntamente com a escolha do conjunto de gestos a ser utilizado e sua organização através de uma taxonomia (que pode utilizar classificações encontradas na literatura, ou mesmo propor novas classificações). Os resultados obtidos nessa etapa são base para a escolha das tecnologias utilizadas nas etapas subsequentes: qual(is) *hardware(s)* será(ão) utilizado(s) na captura dos gestos executados ou na exibição dos resultados obtidos (respectivamente, aquisição de dados e *feedback*), e quais tecnologias de *software* poderão ser adaptadas à necessidade de reconhecimento e classificação dos gestos, com seu vínculo posterior a um evento ou conjunto de eventos de aplicação. Os tópicos a seguir apresentam diferentes tecnologias de *hardware* e *software* utilizáveis no processo de reconhecimento de gestos (cujas etapas podem ser visualizadas na Fig. 3), identificando suas principais características e os pontos positivos e negativos decorrentes da sua adoção.

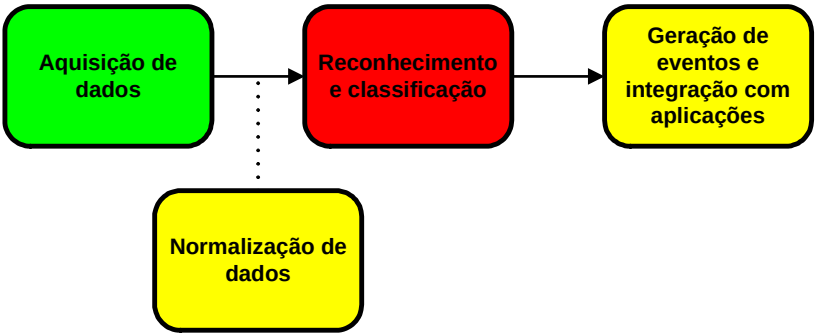

Figura 3: Etapas do processo de reconhecimento de gestos.

3.1 Aquisição de dados

O processo de reconhecimento de gestos tem seu início na fase de aquisição de dados. Essa fase é responsável por coletar as informações que representam os gestos, armazená-las e, opcionalmente, submetê-las a normalizações e filtragens. A normalização visa estabelecer limites (tanto no espaço quanto no tempo) para os gestos executados por diferentes usuários; assim, é possível identificar gestos semelhantes exe-

cutados em janelas de tempo de tamanhos diversos, bem como gestos semelhantes executados de forma mais ou menos esparsa em relação ao espaço. A filtragem, por sua vez, objetiva eliminar ruídos capturados conjuntamente aos dados, os quais podem distorcer os resultados obtidos nas etapas posteriores.

Toda aquisição de dados é baseada em uma tecnologia ou conjunto de tecnologias específico, que por sua vez utiliza um ou mais dispositivos de entrada. Independentemente da tecnologia, é possível classificar os dispositivos utilizados de acordo com uma série de características observáveis (BOWMAN et al., 2004). O número de graus de liberdade, por exemplo, é um indicativo de complexidade que permite relacionar um dispositivo a uma determinada necessidade de interação; a frequência de geração de dados (*discreta* ou *contínua*), juntamente com a especificação da forma como os dados são gerados (*ativa* ou *passiva*), permite que os dispositivos sejam relacionados a posturas ou trajetórias; a *intenção de uso*, por sua vez, permite classificar os dispositivos de acordo com os objetivos pretendidos (localização e seleção de elementos, navegação, entre outros). A adoção desses critérios de classificação, individual ou coletivamente, permite que diferentes dispositivos sejam mapeados para diferentes aplicações através da identificação de funcionalidades presentes no contexto.

Com relação às tecnologias utilizadas para a aquisição de dados gestuais, dois enfoques recebem um maior destaque na literatura: utilização de *luvas instrumentadas e rastreadores de movimento* e técnicas de *visão computacional*. Ambos são apresentados nos próximos tópicos.

3.1.1 Luvas instrumentadas e rastreadores de movimento

As luvas instrumentadas permitem coletar dados que refletem o estado da mão de um usuário em uma determinada posição de uma linha de tempo. Basicamente, os dados que representam o estado em que uma mão se encontra são resumidos aos diferentes graus de flexão que cada um dos dedos pode assumir e à orientação que cada um dos dedos pode assumir em relação à mão como um todo (ou em relação a cada um dos demais dedos). Diferentes modelos de luvas instrumentadas podem ser vistos na Fig. 4.

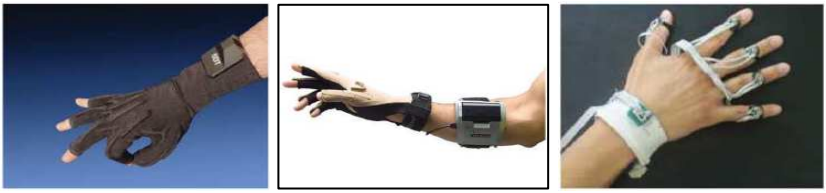

Figura 4: Modelos de luvas instrumentadas (DIPIETRO et al., 2008).

O surgimento dessas luvas, com sua posterior adoção como dispositivos para coleta e mensuração de dados, data da década de 70 (DIPIETRO et al., 2008). Desde então, diversas tecnologias foram desenvolvidas e incorporadas ao *hardware* original, principalmente na parte de sensores (número, disposição e tecnologia de coleta de dados). Considerações sobre a tecnologia de sensores utilizada são importantes pelo fato de determinarem não só as limitações apresentadas pelo dispositivo, mas também os custos de aquisição e manutenção envolvidos.

Luvas instrumentadas são comumente utilizadas na captura de dados referentes a posturas, ou a sequências de posturas. Como vantagens de sua utilização, podem ser citadas as altas taxas de captura (que permitem identificar rápidas trocas de postura executadas pelos usuários), a possibilidade da execução dos gestos com diferentes orientações espaciais e a inexistência de *oclusão* (problema que ocorre quando determinadas partes do corpo ocultam outras partes). A principal desvantagem reside no fato do dispositivo ter de ser vestido; apesar de serem confeccionadas com materiais ajustáveis, as luvas podem não se adaptar a diferentes formatos e tamanhos de mão – o que implica na geração de dados distorcidos (BOWMAN et al., 2004).

Os rastreadores de movimento (também conhecidos como rastreadores de posição), por sua vez, são utilizados na aquisição de dados correspondentes a trajetórias. Essas trajetórias podem ser executadas por diferentes partes do corpo, e se caracterizam pelo posicionamento e pela orientação espacial do movimento, podendo ser delimitadas por um intervalo de tempo pré-estabelecido.

A aquisição dos dados que compõem as trajetórias pode ser realizada através de equipamentos que empregam diferentes tecnologias. Há, porém, restrições genéricas tidas como críticas para a escolha do equipamento adequado (BOWMAN et al., 2004); são elas o *alcance* (distância máxima permitida entre o usuário e o equipamento, ou parte do equipamento), a *latência* (intervalo entre a ocorrência do movimento e a captura do mesmo pelo equipamento), o *ruído* (distorção nos dados

gerados pelo equipamento) e a *precisão* (o quão fidedignos ao movimento real os dados adquiridos realmente são). As diferentes tecnologias de aquisição de dados impõem pesos diferenciados para essas restrições, sendo que a escolha do equipamento a ser utilizado é feita considerando o conjunto de características que melhor se adapta ao contexto do domínio de aplicação.

Dentre as tecnologias existentes, os trabalhos de ROLLAND et al. (2001), WELCH et al. (2002), FOXLIN (2002) e ALLEN et al. (2001) destacam o rastreamento *magnético*, *mecânico*, *acústico*, *inercial*, *óptico* e *híbrido*. No rastreamento magnético, a posição e orientação espacial de um receptor são calculadas em relação a um emissor, responsável pela geração de um campo magnético de baixa frequência. É uma tecnologia precisa, porém susceptível a ruídos gerados pela presença de elementos metálicos no alcance do campo magnético. Rastreadores mecânicos, por sua vez, conectam fisicamente o objeto rastreado a uma base fixa, o que praticamente elimina o ruído gerado pela transmissão *wireless* de dados entre transmissor e receptor. Apesar da alta precisão e baixa latência, esse tipo de rastreador impõe limites à liberdade de movimentos dos usuários, obrigando-os a respeitar o alcance físico da conexão estabelecida. O rastreamento acústico utiliza sons em alta frequência, emitidos a partir de uma fonte e captados por um conjunto de microfones. Duas configurações são características: na primeira, a fonte emissora de sons localiza-se no objeto rastreado, enquanto que os microfones encontram-se dispostos no ambiente (abordagem conhecida como *outside-in*); na segunda, inverte-se o posicionamento da fonte emissora e dos microfones, em uma abordagem conhecida como *inside-out*. É uma tecnologia acessível, porém dependente das características acústicas do ambiente no qual é utilizada, visto que diferentes sons gerados nesse ambiente podem causar interferências e perda de precisão. No rastreamento inercial, giroscópios e acelerômetros são responsáveis pela geração de dados posicionais e de orientação. Usualmente, esses dispositivos são dispostos em um mesmo sensor, o que simplifica a arquitetura do equipamento. Como principal limitação, tanto giroscópios quanto acelerômetros apresentam erros cumulativos, o que pode distorcer os dados adquiridos gerando interpretações errôneas dos gestos executados. No rastreamento óptico, câmeras e sensores são utilizados para captar reflexos ou pulsos luminosos seguindo as abordagens *outside-in* e *inside-out*. Permite um grande número de configurações, que vão desde a determinação do número de câmeras a serem utilizadas até o tipo e o posicionamento dos marcadores que serão rastreados. O ponto falho da tecnologia está na

oclusão, que ocorre quando um ou mais marcadores ficam ocultos e, por conseguinte, não podem ser rastreados. Finalmente, o rastreamento híbrido utiliza diferentes tecnologias em conjunto para obter melhores resultados, através da compensação dos pontos falhos de uma técnica pelos pontos fortes de outra. A maior restrição à sua utilização está na complexidade dos dispositivos, que incorporam duas ou mais tecnologias distintas. Exemplos de rastreadores de movimento podem ser visualizados na Fig. 5.

De acordo com as características apresentadas, a escolha de uma tecnologia para rastreamento de posições deve incluir o ambiente de uso como integrante do domínio de aplicação; dessa forma, não só as necessidades dos usuários e os requisitos a serem atendidos guiarão a escolha, mas também as restrições ambientais que podem interferir diretamente na qualidade dos dados obtidos.

Figura 5: Dispositivo para rastreamento magnético (esquerda)[2] e mecânico (direita)[3].

3.1.2 Técnicas de visão computacional

A aquisição de dados baseada em técnicas de visão computacional utiliza *streams* de vídeo como dados de entrada. Essas técnicas permitem que tanto posturas quanto trajetórias sejam capturadas e, posteriormente, avaliadas. Para posturas, são utilizados *screenshots* ou *frames* escolhidos a partir dos *streams* disponíveis; para trajetórias, *streams* inteiros ou

[2] http://www.inition.co.uk/inition/product.php?URL_=product_mocaptrack_ascension_flockofbirds&SubCatID_=18

[3] http://www.macs.hw.ac.uk/~hamish/9ig2/topic22.html

partes de *streams* com intervalos bem definidos compõem o conjunto de dados.

As técnicas de visão computacional buscam aumentar a naturalidade da utilização de gestos como componentes de uma interface de comunicação humano-computador. O objetivo principal é permitir que o elemento humano interaja livremente com o elemento computacional, sem a necessidade de que o primeiro vista dispositivos instrumentados e fique restrito ao espaço delimitado pelas conexões desses dispositivos ao computador (MITRA et al., 2007). A necessidade de vestir e conectar dispositivos às partes do corpo a serem rastreadas é usada como principal argumento contrário à instrumentação e favorável às técnicas de visão. As limitações impostas por essa necessidade não restringem apenas a naturalidade da utilização dos gestos, mas também a aplicação do conceito de reconhecimento a diferentes áreas como, por exemplo, vigilância (POPPE, 2007).

Para a aquisição dos *streams* de vídeo, diferentes tecnologias podem ser utilizadas. A configuração mais comum baseia-se em *web-cams*, provendo um ambiente de baixo custo e de fácil instalação e organização. O número de câmeras varia de acordo com a necessidade inerente ao domínio de aplicação: gestos executados em 2D podem ser reconhecidos através de *streams* adquiridos por uma única câmera (MANRESA et al., 2005); gestos em 3D, por sua vez, costumam ser reconhecidos a partir de *streams* provenientes de duas ou mais câmeras, combinados por meio de métodos de triangularização (ARGYROS et al., 2006). Outras possibilidades são as câmeras infravermelhas com recurso de *time-of-flight* (BREUER et al., 2007) e câmeras com visão estéreo (GORDON et al., 2008), capazes de gerar padrões representativos de gestos em três dimensões. Nessas câmeras, a terceira dimensão (profundidade) costuma ser codificada através de padrões de cor, permitindo a diferenciação entre elementos mais próximos de elementos mais distantes. Exemplos de câmeras infravermelhas e estéreo podem ser visualizados na Fig. 6.

Com relação ao tratamento dado aos *streams* de vídeo adquiridos, algumas considerações podem ser feitas independentemente da tecnologia de captura utilizada. O reconhecimento de um gesto só poderá ser realizado se a área de interesse (mão, braço, ou outra parte do corpo do usuário) puder ser identificada, destacada e rastreada. Para isso, técnicas como segmentação de imagens (MALIMA et al., 2006), binarização e detecção de contornos (YEUNG et al., 2008) são extensivamente utilizadas. A correta aplicação dessas técnicas (isoladamente ou em conjunto), porém, depende de uma série de fatores ambientais que

podem distorcer os resultados obtidos. Alguns exemplos: gestos executados com planos de fundo heterogêneos podem ser interpretados incorretamente, dada a dificuldade em destacar o objeto de interesse; diferenças na iluminação podem comprometer a segmentação dos objetos de interesse, fundindo-os a outros elementos da cena; e diferentes objetos em movimento podem dificultar o rastreamento do objeto que está executando o gesto.

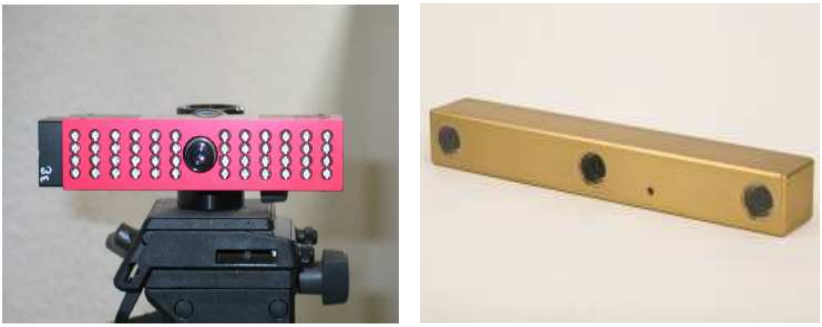

Figura 6: Câmera infravermelha com *time-of-flight* (BREUER et al., 2007), à esquerda, e câmera estéreo (à direita)[4].

Os dados adquiridos por meio de instrumentação ou visão computacional, após devidamente armazenados e processados, compõem as entradas para a(s) técnica(s) de reconhecimento e classificação escolhida(s). As principais idéias que fundamentam essas técnicas são apresentadas a seguir.

3.2 Reconhecimento e classificação

Gestos executados como forma de interação são compostos por um conjunto de características, que podem ser utilizadas isoladamente ou relacionadas a dimensões como espaço e/ou tempo. Independentemente da origem dos mesmos ser instrumentada ou por meio de visão computacional, suas características são extraídas, opcionalmente normalizadas e utilizadas em duas etapas distintas: na primeira, os valores relativos às características relevantes são utilizados como base para a construção ou

[4] http://www.imveurope.com/products/product_details.php?product_id=87

treinamento de classificadores; na segunda, os valores são submetidos aos classificadores previamente construídos, de forma que estes possam relacioná-los a uma das classes (ou padrões) conhecidas. Tem-se um processo claro de reconhecimento de padrões: inicialmente é definida uma representação dos gestos a serem reconhecidos, através da construção ou treinamento de uma estrutura própria para esse fim; essa representação, após devidamente estabelecida, é utilizada no processo decisório subsequente, que classifica (ou não) uma determinada entrada de dados como um padrão gestual válido.

Diversos métodos e algoritmos foram desenvolvidos desde o início dos estudos sobre interação por gestos, cada qual buscando maximizar a acurácia dos resultados de classificação através da minimização dos falsos positivos (reconhecimentos indevidos), ou erros de tipo I, e falsos negativos (reconhecimentos não executados), ou erros de tipo II. Apesar de numerosos, os métodos desenvolvidos podem ser agrupados segundo um conjunto de propriedades similares, que permitem estabelecer taxonomias de classificadores. O relatório técnico escrito por WATSON (1993) classifica os métodos utilizados no reconhecimento de gestos como métodos baseados em *similaridade de modelos*, métodos *conexionistas* e métodos *estatísticos*.

Na similaridade de modelos, os dados adquiridos para classificação são comparados aos padrões previamente conhecidos através de métricas de similaridade, as quais permitem quantificar a semelhança entre os dados e os padrões. É comum a utilização da distância Euclidiana como quantificadora, visto que a mesma pode ser adaptada a espaços *n*-dimensionais (comuns, dependendo das características extraídas dos gestos executados). Os métodos que se enquadram nessa classificação, apesar de apresentarem baixa complexidade de implementação e bons tempos de resposta, não se adaptam adequadamente a entradas de dados heterogêneas – por exemplo, provenientes de diferentes usuários. Pelo fato de não haver treinamento, as estruturas utilizadas apresentam baixa capacidade de generalização, ficando restritas a reconhecimentos baseados em limiares de aceitação. A variação desses limiares se torna, juntamente com o aumento dos dados utilizados na construção dos classificadores, a única forma de melhorar os resultados obtidos.

Os métodos conexionistas, por sua vez, apresentam grande capacidade de generalização, sendo bem representados pelas *redes neurais*. Essas estruturas, após devidamente treinadas, permitem que entradas de dados heterogêneas sujeitas a ruídos, mesmo incompletas, sejam classificadas corretamente. A problemática da utilização desses métodos reside no grande número de parâmetros que podem ser combinados para

formar a estrutura de classificação; assim, diferentes combinações de topologias, funções de ativação, estratégias e taxas de treinamento podem gerar resultados distintos. Além disso, não existem definições estabelecidas sobre como o classificador deve ser construído, restando assim estratégias empiricamente testadas ou mesmo configurações por tentativa e erro.

Os métodos estatísticos buscam utilizar a simplicidade da classificação por similaridade aliada à generalização obtida pelo processo de treinamento. Nesses métodos, os classificadores são treinados de forma a ajustar valores estatísticos de representatividade, os quais são utilizados no momento do reconhecimento como base de comparação. Dessa forma, o conjunto de características extraído do gesto a ser classificado gera valores estatísticos que são comparados com os padrões conhecidos. A maior similaridade entre os valores determina o resultado do reconhecimento. Como exemplos desses métodos, podem ser citados os *modelos ocultos de Markov* (*Hidden Markov Models* - HMM).

O reconhecimento positivo de um gesto, quando identificado, leva a uma decisão por parte da aplicação que utiliza a interface. O tópico a seguir identifica os mapeamentos mais comuns de gestos para eventos e ações executadas pelas aplicações.

3.3 Geração de eventos e integração com aplicações

Após as etapas de aquisição de dados e reconhecimento/classificação, os gestos executados são traduzidos para eventos de aplicação, os quais buscam prover aos usuários o controle necessário sobre o *software* ou o *hardware* ao qual a interface está vinculada. Em termos simples, pode-se dizer que a integração de uma interface baseada em gestos com uma aplicação se dá através do mapeamento de posturas e trajetórias para comandos ou conjuntos de comandos; estes, por sua vez, geram eventos que são interceptados e tratados pela aplicação. Comumente, esses comandos são executados através de teclado e *mouse*.

Os comandos enviados aos *hardwares* e *softwares* a serem controlados são definidos de acordo com o contexto de utilização da aplicação. BOWMAN et al. (2004) os classifica de acordo com os objetivos a serem atingidos. São eles:

- *execução de funções específicas* – através de comandos, os usuários têm condições de invocar funcionalidades disponibilizadas pelas aplicações. Como exemplo, podem ser citadas as

opções de formatação existentes nos *softwares* editores de texto (negrito, itálico, sublinhado, entre outras);

- *alteração do modo de interação* – permite que o comportamento da aplicação seja modificado através da seleção de uma funcionalidade. Pode ser exemplificada pela utilização das opções presentes nas barras de ferramentas, caracterizadas como agrupadoras de funções. A seleção de uma nova ferramenta de trabalho implica na modificação de dados pré-selecionados ou, na ausência destes, na modificação do comportamento da aplicação a partir da escolha da ferramenta;
- *alteração do estado da aplicação* – implica em modificar o contexto de execução da aplicação. Um exemplo que pode ser citado é o da mudança de foco entre diferentes janelas. Cada janela pode estar sendo utilizada com uma finalidade específica; a mudança de foco leva à necessidade de adaptação da aplicação, de forma que a mesma possa responder aos eventos vinculados ao contexto da janela corrente, também chamada de janela de primeiro plano.

O mapeamento de gestos para eventos não segue uma convenção: gestos semelhantes podem ser mapeados para eventos distintos em diferentes aplicações, e vice-versa. Uma prática que visa auxiliar no processo de mapeamento consiste em relacionar a dinamicidade dos gestos à dinamicidade dos eventos. Posturas são utilizadas como geradoras de eventos atômicos, como seleção de itens de menu, seleção de ferramentas, cliques em botões, entre outros. Trajetórias, por outro lado, são relacionadas a eventos contínuos. Caracterizam-se como eventos contínuos aqueles que ocorrem durante um intervalo de tempo, como a movimentação do ponteiro do *mouse* em uma aplicação *desktop* em 2D ou 3D ou a navegação em um ambiente virtual imersivo.

3.4 Considerações sobre o processo de reconhecimento de gestos

Os tópicos anteriores permitem identificar que o processo de reconhecimento de gestos pode ser organizado como um *pipeline*, onde existe uma sequência de execução na qual os resultados obtidos por uma etapa são utilizados como entrada da etapa seguinte. É interessante notar que esse *pipeline* pode ser mais ou menos complexo, dependendo do número de etapas que o compõem; essa composição se dá pelo acrescimo de etapas intermediárias entre a coleta, reconhecimento/classificação

e geração de eventos – como, por exemplo, filtragem e normalização de dados.

Como resultado do processo de reconhecimento de gestos, tem-se um comando ou evento utilizável no controle de uma aplicação. A construção da interface, no entanto, não termina nesse ponto. É necessário validar o trabalho executado, tanto em termos de usabilidade quanto em termos de desempenho. O capítulo a seguir trata da avaliação e *benchmarking* de interfaces, descrevendo seus critérios e etapas.

4. Validação e *benchmarking* de interfaces

A construção de interfaces baseadas em gestos implica em uma validação, cujo foco pode estar centrado tanto no aspecto computacional do processo de reconhecimento quanto no aspecto humano de usabilidade. Essa validação pode ser realizada através de um *benchmarking*, objetivando comparar os resultados obtidos pela utilização do vocabulário proposto com experiências e trabalhos anteriores. A escolha do *benchmarking* a ser aplicado, porém, é dependente do foco computacional ou humano, visto que ambos possuem características distintas passíveis de serem consideradas.

4.1 Foco computacional

Uma validação com foco computacional busca comparar a eficácia e eficiência entre o método proposto com métodos previamente desenvolvidos. Técnicas baseadas em visão computacional podem utilizar como base de *benchmarking* o banco de dados de imagens desenvolvido e disponibilizado por TRIESCH et al. (1996). O banco de dados citado é composto por imagens de dez posturas de mão, executadas por vinte e quatro pessoas em frente a três planos de fundo distintos. Trabalhos como os de FANG et al. (2007), MARCEL (2002) e JUST et al. (2006) utilizam esse banco de dados parcial ou integralmente como conjuntos de treinamento, teste e validação dos novos métodos de classificação propostos. A Fig. 7 apresenta exemplos de imagens disponíveis, com variações nas posturas e nos planos de fundo.

Figura 7: Exemplos de posturas e planos de fundo utilizados para *benchmarking* de métodos baseados em visão computacional[5].

[5] http://www-prima.inrialpes.fr/FGnet/data/09-Pets2002/data/POSTURE/

Outras referências, tanto para posturas quanto para gestos, foram compiladas por MARCEL (2009). Essa compilação relaciona bancos de dados de imagens coloridas e sequências de imagens, para validação do reconhecimento de gestos dinâmicos. A Fig. 8 apresenta uma dessas sequências.

Figura 8: Sequência de imagens representando um gesto dinâmico[6].

Técnicas baseadas em luvas instrumentadas e rastreadores de movimento visam atingir os mesmos objetivos das técnicas baseadas em visão computacional. Porém, nas pesquisas realizadas não foi encontrado um banco de dados de posturas e trajetórias que servisse como base para o *benchmarking* com esses dispositivos. Trabalhos baseados em tecnologias instrumentadas costumam definir um vocabulário próprio para validação, que não necessariamente se assemelha a vocabulários pré-existentes. Como exemplo, o trabalho de WEISSMANN et al. (1999) define um vocabulário composto por vinte posturas de mão, sem a especificação de quais são as vinte posturas escolhidas e qual a base teórica para a escolha. Já o trabalho de ISHIKAWA et al. (1999) especifica seis gestos a serem reconhecidos, baseados no jogo da pedra, papel e tesoura. O autor não especifica se o método será usado em uma interface para uma versão virtual do jogo, e não apresenta outras justificativas para a escolha do vocabulário.

Após a pesquisa realizada sobre métodos de validação com foco no aspecto computacional, é possível identificar as seguintes características pertinentes aos trabalhos selecionados:

- *bases de dados para comparação/validação* – conjuntos de imagens individuais e de sequências de imagens estão disponíveis publicamente para métodos baseados em visão computacional; apesar disso, é comum o desenvolvimento de bases de dados de imagens próprias a cada trabalho, visando complementar as bases de dados comuns. Métodos baseados em luvas instrumentadas e rastreadores de movimento não possuem bases comuns de comparação;

[6] http://www.idiap.ch/resources/gestures/

- *comparação de resultados com trabalhos anteriores* – trabalhos focados na melhoria de técnicas desenvolvidas anteriormente costumam comparar resultados diretamente. Trabalhos que exploram novas abordagens costumam mensurar sua robustez através de variações na configuração do modelo proposto;
- *população utilizada na validação* – é composta por conjuntos de tamanho variável, sendo que cada conjunto pode ser formado por possíveis usuários do método proposto ou por pessoas escolhidas aleatoriamente. Não há especificação para o número de pessoas envolvidas, nem para o volume de testes a ser realizado por cada uma dessas pessoas;
- *métricas* – para ambos os métodos, a validação com foco computacional mensura o número de reconhecimentos positivos e negativos, bem como o número de falsos positivos e falsos negativos, utilizando-os como base para o cálculo de efetividade do processo de reconhecimento.

4.2 Foco humano

A avaliação com foco no fator humano busca analisar, qualificar e testar artefatos componentes de uma interface (ou a interface como um todo), objetivando identificar problemas relacionados à usabilidade e à ergonomia. Segundo BOWMAN et al. (2004), a identificação de tais problemas é o ponto central do processo de avaliação; porém, os resultados também podem ser utilizados como uma forma de entendimento da técnica, dispositivo ou metáfora utilizada. Esse entendimento, por sua vez, pode resultar na definição de guias para o *design* de novas técnicas, dispositivos ou metáforas, servindo como base de conhecimento para o seu desenvolvimento. Outra possibilidade para a utilização dos resultados de uma avaliação é o desenvolvimento de modelos de desempenho, os quais buscam quantificar os resultados de uma combinação formada por usuários, tarefas e interfaces de forma a possibilitar comparações entre diferentes casos de uso.

Na literatura, é possível identificar diferentes classificações para as formas de avaliação relacionadas a interfaces de usuário. Há, porém, alguns termos comuns considerados relevantes nos trabalhos pesquisados; esses termos dividem-se quanto ao tipo de avaliação utilizado (*analítico* ou *empírico*, *formativo* ou *sumativo*) e quanto ao tipo de resultado gerado (*quantitativo* ou *qualitativo*).

No trabalho de HIX et al. (1992), é possível identificar um primeiro nível de classificação relacionado ao momento em que a avaliação de uma interface ocorre. A Fig. 9 exibe a classificação utilizada.

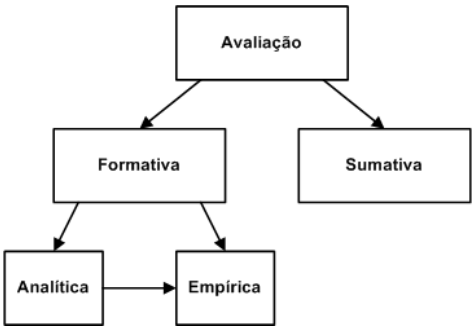

Figura 9: Classificação baseada no modelo *formativo/sumativo*.

Nesse contexto, uma avaliação *sumativa* ocorre após o término do projeto de uma interface, ou durante as etapas finais do projeto. É utilizada muitas vezes como teste de campo para a interface, ou mesmo como uma forma de comparar duas interfaces distintas. A avaliação *formativa*, por sua vez, ocorre iterativamente durante o processo de desenvolvimento da interface. Seu objetivo é corrigir problemas de usabilidade enquanto a interface é construída, através de ciclos de validação bem definidos e distribuídos. O foco dos autores se concentra nesse tipo de avaliação, subdividindo-o em duas modalidades. A primeira delas retrata o modelo *analítico*, através do qual uma interface é avaliada através de métodos formais baseados em projeções de como será o comportamento dos usuários em relação à interface, de acordo com as tarefas a serem executadas. Dada a utilização de projeções, é possível executar a análise antes mesmo da disponibilização de protótipos. Por se basear em modelos formais, assume que a análise é executada por profissionais especialistas em usabilidade. O modelo *empírico*, componente da segunda modalidade, baseia-se na coleta de dados oriundos da observação de usuários representativos e de sua interação com protótipos. O acompanhamento do comportamento dos usuários pode ocorrer de forma controlada (em laboratórios especialmente organizados para tal), ou no próprio local de trabalho dos mesmos, objetivando simular com a maior riqueza de detalhes possível o ambiente e as condições

reais de uso da interface. Apesar de disjuntos por definição, os dois modelos se completam pelo fato do empirismo ser considerado um validador para a análise.

O trabalho de HARTSON et al. (2003), por sua vez, adota duas possibilidades de classificação para um primeiro nível de análise: uma delas semelhante à classificação de HIX et al. (1992), e outra considerando diretamente o modelo analítico/empírico. Como um complemento à definição dada anteriormente, os modelos analítico e empírico são relacionados, respectivamente, a especialistas em usabilidade e a usuários representativos. Essa relação permite identificar que a diferença do público-alvo das validações (especialistas e usuários) implica em uma escolha de métodos coerentes a cada um.

MAZZA (2006) assume que o primeiro nível de classificação utilizado é o dos modelos analítico/empírico, sendo que o último é dividido quanto ao tipo de resultado gerado. A Fig. 10 exibe a classificação proposta.

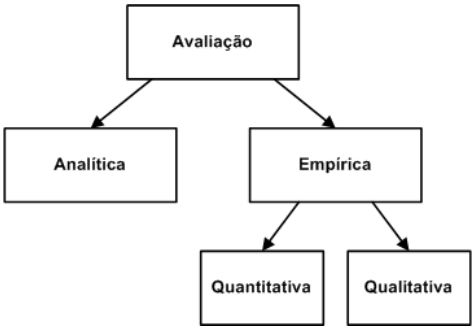

Figura 10: Classificação baseada no modelo *analítico/empírico*.

Para o autor, o modelo analítico é baseado em modelos humanos de raciocínio e comportamento, enquanto que o modelo empírico (também conhecido como *estudo de usuário*) trata diretamente com dados obtidos através de usuários reais. No modelo empírico, consideram-se dois tipos de resultados ou *estudos*: quantitativos, obtidos através de medições feitas sobre hipóteses de uso (como, por exemplo, o desempenho de um usuário em relação a uma tarefa específica ou o número de tentativas necessário para a completude de uma tarefa específica), e qualitativos, obtidos a partir da análise de opiniões e comportamentos dos usuários durante seções de teste e validação.

Apesar das diferentes classificações apresentadas na literatura, a primeira etapa para a definição e escolha de métodos de avaliação é o conhecimento das ferramentas disponíveis e que sejam reconhecidamente úteis para a tarefa. Existem ferramentas que podem ser utilizadas independentemente da complexidade da interface a ser avaliada; dentre elas, destacam-se:

- *análise de tarefas dos usuários* – corresponde ao popular levantamento de requisitos e modelagem de casos de uso da engenharia de software. Permite identificar quais são as tarefas executadas pelos usuários na aplicação, bem como relacionar ações atômicas ou sequenciais, descrever relacionamentos entre ações e diagramar o fluxo de informações correspondente às tarefas. Depende massivamente das informações fornecidas pela população de usuários representativos, bem como da natureza das tarefas executadas e das necessidades organizacionais identificadas (como pode ser observado na Fig. 11);

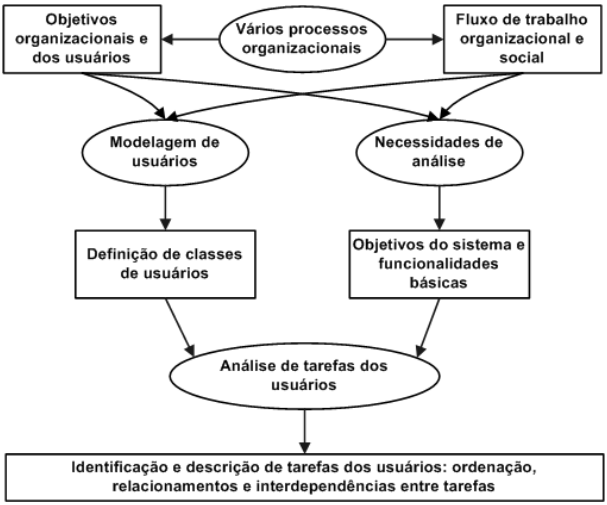

Figura 11: Contexto de aplicação da análise de tarefas dos usuários (GABBARD et al., 1999) – adaptação.

- *cenários* – permitem visualizar o fluxo de trabalho dos usuários, utilizando informações previamente coletadas pela análise das tarefas dos mesmos. A sua correta construção permite que mé-

todos de avaliação posteriores possam identificar problemas de usabilidade, e também avaliar soluções alternativas a situações não previstas relativas às tarefas a serem executadas;

- *taxonomias* – através da classificação de tarefas, é possível identificar similaridades entre possíveis ações e métodos de avaliação. Também é possível estabelecer divisões de tarefas em tarefas menores, de forma a permitir identificar técnicas avaliativas específicas para as tarefas resultantes dessas divisões. Isso permite compor uma avaliação adequada a tarefas de nível mais alto através da união de técnicas relacionadas a tarefas de nível mais baixo;
- *protótipos* – oferecem um resultado (mesmo que prévio) da modelagem da interface, permitindo aos usuários validarem o modelo e relatarem suas experiências. Podem ser utilizados em diferentes fases do processo de desenvolvimento, não precisando estar completos em termos de projeto e funcionalidade. Um exemplo de aplicação para protótipos é descrito pelo paradigma *Oz*, também conhecido como paradigma *Wizard of Oz* (KELLEY, 1984). O paradigma se baseia na monitoração dos testes feitos pelos usuários, sem que estes necessariamente saibam que estão sendo monitorados. Durante a monitoração, é possível ao profissional que acompanha os testes assumir o papel da interface, fornecendo *feedback* aos usuários e acompanhando o comportamento dos mesmos.

É importante ressaltar que a utilização das ferramentas citadas anteriormente objetiva a coleta de informações acerca de quais tarefas devem ser executadas e como essas tarefas devem ser executadas. O raciocínio sobre como isso pode ser melhorado começa a tomar forma a partir da aplicação de técnicas de avaliação mais complexas, fundamentadas na utilização de uma ou mais dessas ferramentas. Uma compilação de técnicas de avaliação, que podem ser aplicadas a diferentes tipos de interface foi executada por BOWMAN et al. (2004), e resultou no seguinte conjunto: *acompanhamento cognitivo*, *avaliação heurística*, *questionários* e *entrevistas/demonstrações*.

O acompanhamento cognitivo busca avaliar o comportamento de uma interface em relação a usuários pouco frequentes ou que estejam utilizando a interface pela primeira vez. Para a avaliação são consideradas tarefas que serão comumente utilizadas, sendo que cada etapa dessas tarefas é avaliada. Como resultado direto é possível

observar o quão intuitiva é a interface pela facilidade de entendimento demonstrada pelos usuários.

As opiniões sobre a maturidade de uma interface nem sempre partem dos usuários finais da aplicação; às vezes, a avaliação é direcionada a especialistas em usabilidade e ergonomia. Na avaliação heurística, esses profissionais são responsáveis por aplicar seus conhecimentos buscando encontrar pontos falhos no projeto e implementação de uma interface em particular. Os resultados obtidos são classificados por prioridade, e as correções necessárias são feitas de forma iterativa. Não há a participação de usuários representativos em nenhum momento; portanto, a avaliação pode não considerar aspectos pontuais decorrentes da utilização constante e do processo adaptativo relacionado a situações imprevistas.

As técnicas citadas anteriormente podem se utilizar da aplicação de questionários ou da adoção de entrevistas e demonstrações para a coleta e o registro de resultados. Por serem escritos e poderem ser respondidos em momentos diversos, os questionários apresentam a característica da flexibilidade e oferecem aos usuários a liberdade de escolher o momento e o local para respondê-los. Porém, o alto grau de formalismo inerente à sua aplicação implica em respostas também formalizadas, destituídas de duplo significado e coerentes de forma a permitir tabulação. Sua importância pode ser observada em trabalhos como o de RIZZO et al. (2005), no qual as diferenças de sexo, habilidades de percepção visual e grau de escolaridade têm relevância na avaliação de interfaces. Essas informações foram obtidas através de questionários aplicados junto à população de interesse, permitindo vislumbrar que além dos quesitos de ergonomia e usabilidade normalmente utilizados, fatores pessoais contribuem na composição das métricas de validação. Por sua vez, as entrevistas e demonstrações, diferentemente dos questionários, admitem um baixo grau de formalismo e completa adaptabilidade: é permitido ao entrevistador conduzir a entrevista de forma a obter os resultados desejados, mesmo que para isso as questões originalmente propostas sejam descartadas e novas questões sejam feitas. Além disso, entrevistadores experientes podem identificar nuances subjetivas relacionadas às respostas dadas, o que enriquece o conteúdo qualitativo que pode ser obtido. Em conjunto às questões apresentadas, protótipos de interfaces podem ser apresentados e utilizados como base para discussão.

Apesar de bem-definidas individualmente, as técnicas de avaliação citadas não costumam ser utilizadas de forma isolada. Diferentes combinações dessas técnicas, com uma sequencialidade especificada *a*

priori, permitem que os resultados gerados por uma técnica sejam utilizados como informações de entrada por outra técnica. Essa abordagem é utilizada como uma forma de garantir que questões inerentes à usabilidade sejam primeiramente respondidas por especialistas, deixando detalhes mais específicos correspondentes à utilização contínua da interface para a avaliação por usuários representativos. Como resultado, a utilização de combinações de técnicas garante uma evolução natural do processo de avaliação, possibilitando inclusive uma redução de custos. Uma sugestão de sequencialidade constituída pelos métodos vistos anteriormente pode ser observada na Fig. 12. Nesta figura, é possível identificar uma sequência de métodos de avaliação iniciada pela avaliação heurística, seguida pela avaliação formativa e finalizada pela avaliação sumativa. A avaliação heurística e a avaliação formativa podem ser executadas *n* vezes, indicando uma iteratividade necessária ao refinamento da interface. É possível identificar também três características derivadas da utilização dessa sequência em específico:

- *custo* – é menor na fase de avaliação heurística pelo fato do número de profissionais necessário ser reduzido, e também por esses profissionais já estarem disponíveis para a modelagem da interface. Conforme os usuários representativos são envolvidos (na avaliação formativa/sumativa), o custo aumenta devido à quantidade de pessoas e à necessidade de alocação de horas das mesmas para as validações;
- *generalidade* – avaliações heurísticas costumam ser genéricas, abrangendo a interface como um todo. As demais avaliações da sequência são mais específicas e, quanto maior o número de usuários envolvidos, maior a especificidade resultante;
- *precisão* – por utilizar a opinião dos usuários, a avaliação sumativa costuma ser extremamente precisa no relato de resultados; já a avaliação heurística oferece indicações mais genéricas das situações de erro, não identificando soluções diretas para os problemas de usabilidade.

O trabalho de GABBARD et al. (2003) exemplifica a utilização da sequência de avaliações exposta na Fig. 12. Esse trabalho apresenta os resultados da avaliação de três interfaces distintas, sendo que duas situações são abordadas: duas das interfaces são baseadas em *software* e uma é baseada em *hardware*. Apesar dessa diferença, o trabalho relata o sucesso na aplicação da técnica, deixando claro que múltiplas iterações são necessárias para um refinamento adequado da interface. Todas as

iterações ocorrem na primeira e na segunda etapas da sequência (de menor custo), gerando protótipos de maior qualidade para a avaliação sumativa (de maior custo).

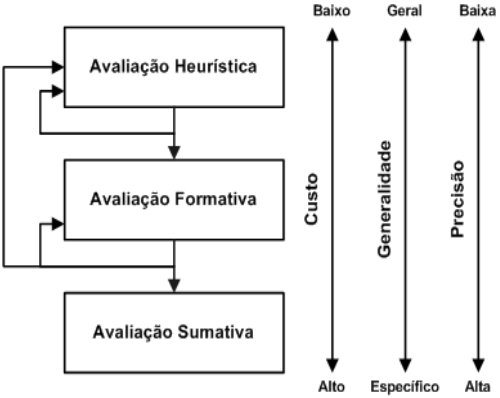

Figura 12: Sequência de aplicação de técnicas de validação (GABBARD et al., 1999) – adaptação.

De acordo com o exposto nos parágrafos anteriores, algumas variáveis podem ser destacadas como determinantes em um processo de classificação de métodos e técnicas de avaliação de interfaces:

- *foco*: computacional ou humano (métricas ou usabilidade);
- *momento*: durante ou após o projeto da interface (formativo/sumativo);
- *população*: profissionais especialistas em usabilidade ou usuários comuns (analítico/empírico);
- *resultados*: indicativos de desempenho e opiniões (dados quantitativos e qualitativos).

É possível aos profissionais responsáveis por processos de validação combinar essas variáveis entre si, escolhendo métodos e dispondo-os de forma a atingir os melhores resultados possíveis para um determinado caso. Essa flexibilidade, porém, dificulta uma classificação unificada dos métodos existentes. É possível, por exemplo, incluir usuários representativos na fase de validação formativa. A vantagem dessa abordagem está na obtenção de conhecimento empírico em diferentes etapas do desenvolvimento da interface; a desvantagem está nos custos envolvidos para a alocação dos usuários nas etapas iterativas

necessárias. Também é possível adotar entrevistas e questionários apenas na fase de validação sumativa, considerando como público-alvo a população de usuários representativos. Essa abordagem ignora os resultados qualitativos que podem ser obtidos com a população de especialistas em usabilidade, durante a fase de validação formativa. Em suma, as combinações possíveis auxiliam pela diversidade, mas complicam na definição de um padrão unilateral de classificação.

Os capítulos dois, três e quatro apresentaram os fundamentos necessários ao entendimento dos componentes de uma interface gestual, sua organização e como se dá a avaliação de uma interface em termos de métricas humanas e computacionais. O capítulo cinco (a seguir) lista um conjunto de trabalhos relacionados, cuja estrutura se enquadra nas definições apresentadas anteriormente. O conteúdo deste capítulo fundamenta a escolha dos métodos cuja precisão e desempenho são comparados neste trabalho.

5. Trabalhos relacionados

A busca por melhores resultados no processo de reconhecimento de gestos leva ao desenvolvimento de novas estratégias, bem como ao aperfeiçoamento de técnicas já consolidadas. Na literatura relacionada, um grande número de trabalhos busca relatar experiências de criação e configuração de diferentes métodos, juntamente com os resultados obtidos. Este capítulo apresenta um conjunto de trabalhos correlatos, cujo objetivo comum é melhorar a precisão de diferentes técnicas. Os trabalhos relacionados são agrupados de acordo com as estratégias de reconhecimento utilizadas, e servem como base para a escolha dos métodos que serão avaliados e comparados nesta dissertação.

5.1 Redes neurais

As redes neurais são utilizadas tanto no reconhecimento de posturas quanto no reconhecimento de trajetórias, dada a sua capacidade de generalização. No trabalho desenvolvido por STERGIOPOULOU et al. (2009), uma rede do tipo SGONG (*Self-Growing and Self-Organized Neural Gas*) foi utilizada em conjunto a um método probabilístico para classificar posturas. O método descrito pelos autores divide-se em quatro etapas: detecção da região da mão do usuário com base na cor da pele do mesmo (o que é feito através de segmentação das imagens usadas como entrada, considerando o espaço de cor *YCbCr*), extração de dados referentes à morfologia da mão do usuário através da utilização da rede neural, identificação dos dedos flexionados e estendidos (base utilizada para a definição dos padrões de gestos) e reconhecimento (baseado em um método probabilístico que considera as combinações possíveis das posições dos dedos do usuário). Para que possa ser aplicado, o método impõe algumas restrições: os gestos devem ser executados pelos usuários utilizando a mão direita, com o braço na posição vertical, palma da mão voltada para a câmera e plano de fundo uniforme. Com a observância das restrições, o método atingiu uma taxa de reconhecimento de 90,45%, em um conjunto formado por 31 diferentes classes de gestos. O trabalho de BAILADOR et al. (2007), por sua vez, utiliza redes neurais como classificadoras de sinais provenientes de acelerômetros. São utilizadas redes do tipo CTRNN (*Continuous Time Recurrent Neural Network*), com a finalidade de prever as acelerações medidas nos eixos *X*, *Y* e *Z* com base nas acelerações atuais geradas durante a execução de trajetórias. A estratégia proposta

estabelece que o menor erro de predição oriente o reconhecimento; assim, a rede que conseguir prever as futuras acelerações com o menor erro determina a classe do gesto reconhecido. São utilizadas oito redes, cada uma representando uma classe de gesto conhecida; as trajetórias dos gestos podem ser visualizadas na Fig. 13. Em um ambiente controlado, com o usuário executando apenas os gestos conhecidos e adotando um protocolo de execução bem definido (com pausas entre os gestos), o método proposto atingiu uma taxa de reconhecimento de 94%; em um ambiente normal, onde a execução dos gestos foi feita de forma intercalada com outras atividades cotidianas, a taxa de reconhecimento decaiu para 63,6%. É importante ressaltar que tanto os dados de treinamento quanto os dados de teste foram coletados de um indivíduo, apenas.

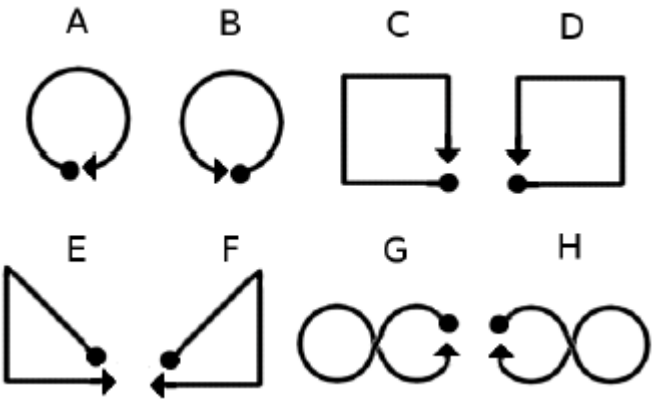

Figura 13: Trajetórias dos gestos utilizados para validação do método baseado em CTRNN (BAILADOR et al., 2007).

Dados instrumentados foram utilizados por XU et al. (2006) como entrada para uma rede neural do tipo *feed-forward*, treinada com o método *backpropagation*. O reconhecimento de posturas, bem como de rotações e translações das mãos dos usuários, se dá em um ambiente virtual de treinamento para uso de armas automotoras. A rede neural opera com 18 entradas (adquiridas de uma *data glove*), uma camada oculta e 15 saídas (uma para cada classe de gesto conhecida). Cinco pessoas contribuíram com os testes, gerando um conjunto de treinamento com 200 instâncias e um conjunto de testes com 100

instâncias. Após a avaliação, foi obtida uma taxa de reconhecimento de 98%.

5.2 Support Vector Machines (SVM)

REN et al. (2009) propõe a utilização de SVMs combinadas a uma técnica conhecida como MEB (*Minimum Enclosing Ball*) para a classificação de posturas. A combinação, batizada de MEB-SVM, objetiva reduzir o tempo e o esforço computacional do reconhecimento, permitindo a utilização de vocabulários mais complexos. Como dados de entrada, são utilizadas imagens cujo tratamento inclui segmentação, binarização e seleção de contornos. Os dados resultantes são utilizados como conjuntos de treinamento e testes, resultando em uma acurácia média de 92,89%. Representações compactas de dados relativos a movimentos foram utilizadas como entrada para o método proposto por MENG et al. (2007), que utiliza uma SVM linear como classificador. Através da subtração de frames de vídeo, os movimentos foram isolados e representados de forma compacta utilizando as técnicas de *Motion History Image* (MHI) e *Hierarchical Motion History Histogram* (HMHH), objetivando viabilizar o reconhecimento em tempo real. Após essa etapa de pré-processamento, os dados obtidos são utilizados como conjuntos de treinamento e avaliação da SVM. Os experimentos realizados para validação do método consideraram seis classes de movimentos, executados com planos de fundo variados por 25 voluntários. A taxa máxima de reconhecimento atingiu 93,1%. O trabalho de CHEN et al. (2007) propõe a utilização de uma SVM para o reconhecimento de posturas, cujos dados são provenientes de *webcams*, no contexto de jogos como o popular *joquempô* (pedra, papel e tesoura). O método desenvolvido no trabalho objetiva prover robustez ao reconhecimento de posturas executadas em diferentes ângulos e por pessoas com diferentes tonalidades de pele. Para isso, as imagens adquiridas são redimensionadas e convertidas para tons de cinza; em seguida, seus histogramas são normalizados e utilizados como entrada da SVM (tanto para treinamento quanto para avaliação). A proposta obteve um reconhecimento positivo de 95%, considerando treinamento e testes executados com dados da mão direita dos usuários. Quando submetido a treinamento com dados de mão direita e testes com dados de mão esquerda, a proposta atingiu um reconhecimento máximo de 90%. Exemplos das posturas utilizadas podem ser visualizados na Fig. 14. LIU et al. (2008) apresenta, em seu trabalho, uma combinação de

SVM e uma técnica de visão computacional conhecida como *Hu moments*, que permite representar imagens através de características independentemente de sua translação, escala e rotação. A proposta visa automatizar a verificação da integridade das mãos dos candidatos à licença para dirigir, na China. As imagens adquiridas são convertidas para o espaço de cor *YCbCr*, segmentadas e suas características extraídas e utilizadas como dados de treinamento para a SVM. Testes com dados de 20 voluntários obtiveram uma acurácia de 96,5%.

Figura 14: Exemplos de imagens de posturas (CHEN et al., 2007).

5.3 Modelos de Markov (Markov Models – MM) e Modelos Ocultos de Markov (Hidden Markov Models – HMM)

Gestos dinâmicos são caracterizados pela representação seriada de pontos compondo trajetórias, as quais apresentam variações espaciais e temporais de execução. Os modelos de Markov e os modelos ocultos de Markov são extensivamente usados no reconhecimento deste tipo de gesto, e são exemplificados nos trabalhos a seguir. CARIDAKIS et al. (2008) utilizou modelos de Markov para representar os aspectos temporais das trajetórias vinculadas aos gestos, combinadas a mapas auto-organizáveis responsáveis pelos aspectos espaciais das mesmas. Após uma fase de normalização dos dados adquiridos, o método proposto atingiu uma acurácia de 93% para um universo de 30 classes de gestos. O trabalho de ELMEZAIN et al. (2008) utilizou dados adquiridos por uma câmera estéreo, objetivando reconhecer as trajetórias vinculadas aos gestos. O método proposto utilizou 36 classes de gestos, sendo 26 letras (A-Z) e 10 números (0-9); alguns exemplos de gestos sendo executados podem ser observados na Fig. 15. Para a classificação, foi utilizado um HMM configurado com a topologia LRB

(*Left-Right Banded*), composto por nove estados e treinado com o algoritmo *Baum-Welch*. Uma taxa de reconhecimento de 94,72% foi obtida durante a avaliação da proposta.

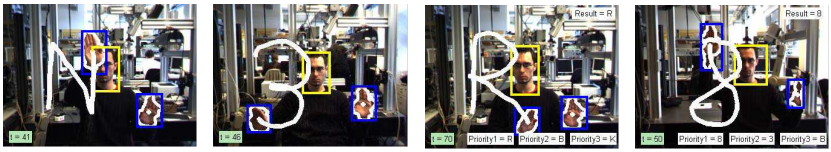

Figura 15: Exemplos de imagens de trajetórias (ELMEZAIN et al., 2008).

Dados instrumentados foram adquiridos por SCHLÖMER et al. (2008) a partir de um *Wiimote*, *joystick* usado pelo console Wii™ da Nintendo®. As acelerações captadas pelo joystick nos eixos *X*, *Y* e *Z*, após filtradas e quantizadas, serviram como base de treinamento e avaliação para um conjunto de cinco gestos. Utilizando um HMM com oito estados, foram atingidas taxas de reconhecimentos positivos entre 85% e 95%. FROLOV et al. (2008), em seu trabalho voltado a *feedback* háptico multimodal, utilizou dados instrumentados para reconhecimento de posturas e gestos. Com uma luva instrumentada, adaptada de forma a prover dados relativos à temperatura e tato, uma taxa de reconhecimento direto de 84% foi atingida (utilizando uma rede de HMMs como classificadores). Considerando o contexto de aplicação, e minimizando um conjunto de erros não relacionados ao *feedback* háptico multimodal, a taxa final de reconhecimento foi considerada pelos autores como sendo de 99%.

5.4 Outras técnicas baseadas em reconhecimento de padrões

Além dos métodos citados anteriormente, tidos como clássicos no contexto de reconhecimento de gestos, diferentes adaptações de métodos utilizados no reconhecimento de padrões em outras áreas de aplicação costumam ser integradas ao conjunto de soluções possíveis para o problema. É o caso, por exemplo, do trabalho de DELLER et al. (2006), que utiliza dados instrumentados para reconhecer posturas e rotações do pulso dos usuários. A técnica de classificação utilizada é baseada no cálculo de distâncias entre a sequência de gestos executada e os gestos armazenados como padrões a serem reconhecidos. Não existe uma etapa

de aprendizado propriamente dita; assim, as instâncias de treinamento são utilizadas diretamente no processo de reconhecimento. Caso a distância entre o gesto executado e um dos padrões conhecidos esteja dentro de um limite previamente estabelecido, o reconhecimento é positivo. Se os gestos reconhecidos estiverem na sequência correta (determinada no momento da gravação), o evento correspondente à sequência é gerado. STEFAN et al. (2008), por sua vez, propõe um método de classificação baseado em *nearest neighbor*, executado sobre dados provenientes de trajetórias. O objetivo principal é reduzir os problemas relacionados às variações de escala e translação inerentes aos gestos, de forma a reconhecê-los e classificá-los independentemente da distância entre o usuário e a câmera e a posição assumida pelo usuário para executá-los. Quando aplicado ao reconhecimento de dígitos entre zero e nove, o método atingiu uma taxa de reconhecimento de 96,3%; essa taxa permaneceu constante independentemente das variações de escala e translação existentes, permitindo uma maior flexibilidade de movimento aos usuários. O trabalho de TARRATACA et al. (2009) utiliza uma combinação da técnica de *nearest neighbor* com HMM para reconhecer e classificar, respectivamente, posturas e gestos (sendo os gestos compostos por uma sequência de posturas). As etapas de aquisição de dados (imagens), pré-processamento e reconhecimento são executadas em dispositivos móveis, atingindo uma acurácia média de 83,3% em um universo de três gestos, cada qual composto por cinco posturas. ZIAIE et al. (2008) propõe a utilização de duas técnicas combinadas: *K-nearest neighbor* e um classificador Bayesiano. O trabalho objetiva classificar posturas pertencentes a três classes distintas, utilizadas no contexto do projeto JAST – *Joint Action Science and Technology*, cujo foco se concentra nos aspectos cognitivos e comunicativos de agentes que operam em conjunto, sejam eles humanos ou artificiais. A integração entre os agentes ocorre por meio do reconhecimento de gestos executados pelo agente humano, que são mimetizados pelo robô (o agente artificial). Utilizando invariantes geométricos como dados de entrada, obtidos pela segmentação e seleção de contornos das mãos dos usuários, o método atingiu uma acurácia de cerca de 93% em um tempo de 50 milissegundos por classificação. Outro trabalho voltado à interação humano-robô foi desenvolvido por KJELLSTRÖM et al. (2008), objetivando mapear o ato de agarrar objetos. No trabalho, frames extraídos de *streams* de vídeo são utilizados como dados de entrada, permitindo a classificação dos gestos de acordo com uma taxonomia. A classificação é executada pela técnica de *K-nearest neighbor*, melhorada por uma técnica de aproximação

conhecida como *Locality Sensitive Hashing* (LSH). Os resultados obtidos pela aplicação das técnicas foram semelhantes aos resultados obtidos através da classificação executada por humanos – o equivalente a 75% de reconhecimentos positivos.

5.5 Considerações sobre os trabalhos relacionados

Através da avaliação dos trabalhos relacionados nos itens anteriores, é possível identificar algumas características comuns. Inicialmente, percebe-se que cada trabalho adota uma técnica específica, implementando-a através de um algoritmo ou da junção de dois ou mais algoritmos, que se complementam em termos de funcionalidade. É característico à grande maioria dos trabalhos, também, a adoção de conjuntos de dados de treinamento e avaliação definidos e construídos de maneira *ad hoc*, sem um embasamento feito em trabalhos anteriores ou mesmo em uma justificativa inerente ao contexto. Finalmente, percebe-se que o objetivo principal dos trabalhos é quantificar o processo de reconhecimento de gestos, através da avaliação da precisão das técnicas abordadas em termos de reconhecimentos positivos.

Embora os trabalhos apresentem resultados expressivos em termos de acurácia, as características comuns identificadas dificultam a comparação entre as técnicas abordadas e, em última instância, a escolha de um método a ser efetivamente adotado na construção de uma interface gestual. A falta de comparações entre métodos, unida à utilização de dados de treinamento e validação discrepantes, impossibilita uma avaliação direta. Essa impossibilidade pode ser retratada por perguntas como: *"a acurácia se manterá, caso os dados de treinamento sejam substituídos pelos dados utilizados no trabalho de fulano"* ou *"como o método se comportará, caso os algoritmos sejam parametrizados de acordo com o trabalho de beltrano"*? Há, ainda, o quesito desempenho. Nem todos os trabalhos explicitam os tempos necessários à construção e treinamento dos classificadores utilizados, bem como à avaliação dos gestos executados. Essa métrica desempenha um papel essencial no processo de escolha de um determinado método, pois é através dela que se identifica a possibilidade da sua utilização em interfaces de tempo real.

Considerando as limitações citadas, a presente dissertação busca comparar um conjunto de métodos utilizados no processo de reconhecimento de posturas e trajetórias, de forma a identificar as melhores técnicas. Como diferencial aos trabalhos supracitados, os métodos esco-

lhidos são comparados entre si em termos de precisão e desempenho, identificando sua adaptação e generalização a um conjunto de dados comum e sua aplicabilidade em interfaces com restrições temporais. O conjunto de dados de treinamento e avaliação é tratado como um subproduto da dissertação, e é disponibilizado publicamente para *download*[7]; dessa forma, outros autores interessados em validar suas técnicas e contribuições têm a opção de utilizar esses dados como ponto de partida, inclusive comparando os seus resultados com os resultados apresentados neste trabalho.

Para que a comparação entre os métodos escolhidos possa ser executada, é necessária a construção de um ambiente composto por elementos de *hardware* e *software* e um vocabulário de gestos. O capítulo seis descreve esse ambiente em detalhes.

[7] http://www.lapix.ufsc.br/index.php/gil-gesture-interaction-layer

6. Ambiente experimental

A comparação entre técnicas de reconhecimento de gestos demanda aquisição de dados e reconhecimento/classificação, duas das etapas citadas e explicadas no capítulo três. No presente capítulo, essas etapas são especificadas em termos de *hardware* e *software*, sendo complementadas pelo vocabulário comum definido como base de treinamento e avaliação dos métodos escolhidos.

6.1 Especificação de hardware

Neste trabalho foram utilizados dados instrumentados, tanto para as posturas quanto para as trajetórias. Essa escolha se deu pela disponibilidade do *hardware* necessário, e pela motivação relacionada à criação de um vocabulário instrumentado de gestos, passível de ser disponibilizado publicamente a outros pesquisadores.

Os dados referentes às posturas foram adquiridos através de uma *data glove* modelo *5DT Data Glove 5 Ultra*[8], fabricada pela empresa *Fifth Dimension Technologies*[9]. A luva utilizada adapta-se à mão direita dos usuários, e é dotada de cinco sensores de fibra óptica – um para cada dedo – objetivando mensurar seus graus de flexão. A Tabela 1 relaciona um conjunto de especificações mais detalhado.

Tabela 1: Especificações da luva instrumentada utilizada no presente trabalho.

Item	**Especificação**
Material	*Lycra*
Sensores	- Fibra óptica
	- 1 sensor para cada dedo, mensurando seu grau de flexão
Interface	- USB 1.1
	- RS-232 (*kit* opcional)
Alimentação	Via interface
Taxa de coleta de dados	75Hz (mínimo)

Fonte: *5DT Data Glove 5 Ultra Series*[10] – adaptação.

O acesso aos dados gerados pela luva é possível através de uma API (*Application Programming Interface*), que disponibiliza funções para conexão, calibragem, recuperação de valores dos sensores (indi-

[8] http://www.5dt.com/products/pdataglove5u.html
[9] http://www.5dt.com/
[10] http://www.5dt.com/downloads/dataglove/ultra/5DTDataGloveUltraDatasheet.pdf

vidual ou coletivamente), entre outras. A API permite a recuperação de dois tipos de dados: *raw* (dados adquiridos diretamente dos sensores) e *scaled* (dados normalizados). A diferença entre ambos está no intervalo de valores gerado: para dados *raw*, os valores adquiridos encontram-se no intervalo [0, 4.095], enquanto que para dados *scaled* os valores encontram-se no intervalo [0, 1]. Quanto maior o valor adquirido, maior é o grau de flexão do dedo relacionado ao sensor. O Quadro 2 apresenta uma postura em termos de valores obtidos a partir dos sensores citados anteriormente.

Quadro 2: Postura representada por valores instrumentados.

Sensores		**Dados**	
Posição (dedo)	**Índice para o *driver***	***Raw***	***Scaled***
Polegar	0	3.807	0,149635
	1	3.807	0,149635
	2	0	0
Indicador	3	3.086	0,964564
	4	3.086	0,964564
	5	0	0
Médio	6	3.447	0,729363
	7	3.447	0,729363
	8	0	0
Anelar	9	2.702	0,69604
	10	2.702	0,69604
	11	0	0
Mínimo	12	2.148	0,145
	13	2.148	0,145
	14	0	0
	15	0	0
Ângulo de inclinação	16	2.048	0
Ângulo de rotação	17	2.048	0

Apesar de a luva utilizada possuir apenas cinco sensores, o *driver* de comunicação acessado pela API disponibiliza um *frame* de dados formado por 18 elementos. Essa estratégia objetiva prover apenas um *driver* e uma API para diferentes modelos de luva, simplificando a arquitetura e a integração de dispositivos heterogêneos. Neste trabalho

são utilizados os elementos correspondentes aos índices *0*, *3*, *6*, *9* e *12* (em destaque no Quadro 2), ignorando os demais.

As trajetórias, por sua vez, são representadas por conjuntos de pontos em um espaço tridimensional. Esses pontos são adquiridos através de um rastreador de movimento baseado em tecnologia magnética, fabricado pela empresa *Ascension Technology Corporation*[11], modelo *Flock of Birds*®[12]. O dispositivo utiliza um campo magnético para rastrear a posição e orientação de um sensor com seis DOF. Maiores detalhes sobre o equipamento podem ser visualizados na Tabela 2.

Tabela 2: Especificações do rastreador de movimento utilizado no presente trabalho.

Item	**Especificação**
Alcance de rastreamento	0,75m (maior acurácia) 0,90m (menor acurácia)
Dados gerados	Coordenadas posicionais nos eixos *X*, *Y* e *Z* Ângulos de rotação nos eixos *X*, *Y* e *Z* Matrizes de rotação *Quaternions*
Interface	RS-232 (velocidade máxima de 115.200 *baud*
Taxa de coleta de dados	Até 144 aquisições/segundo

Fonte: *Flock of Birds – Real-time motion Tracking*[13] – adaptação.

De forma similar à luva instrumentada, os dados do rastreador de movimento são recuperados utilizando-se um conjunto de funções disponibilizadas por uma API. O *frame* de dados recuperado pode ser composto por informações posicionais, angulares ou uma combinação de ambas, de acordo com o modo de operação configurado no dispositivo. Neste trabalho, apenas as informações posicionais – coordenadas dos eixos *X*, *Y* e *Z* – são utilizadas; as demais informações são descartadas.

Os dados apresentados no Quadro 3 descrevem uma trajetória de acordo com os *frames* adquiridos através do rastreador de movimento. No exemplo, a trajetória é formada por 128 pontos em um espaço tridimensional, cada qual podendo ser localizado pelas suas coordenadas *X*, *Y* e *Z*. Como informação adicional, o *timestamp* de aquisição dos dados referentes a cada ponto (em milissegundos), relativo ao início da

[11] http://www.ascension-tech.com/
[12] http://www.ascension-tech.com/realtime/RTflockofBIRDS.php
[13] http://www.ascension-tech.com/docs/Flock_of_Birds.pdf

captura dos *frames* de dados, é armazenado juntamente com as coordenadas. A extensão de uma trajetória pode ser computada pelo número de pontos que a forma, e sua duração pode ser calculada pela diferença dos *timestamps* inicial e final. Tanto a extensão quanto a duração de uma trajetória variam de acordo com a complexidade do gesto relacionado, sua velocidade de execução e o comportamento do usuário que está executando o gesto.

Quadro 3: Trajetória representada por valores instrumentados.

Índice do ponto na trajetória	**Coordenadas**			***Timestamp***
	X	**Y**	**Z**	
0	13.440	4.064	-14.140	62
1	13.456	4.092	-14.148	109
2	13.556	4.208	-14.116	156
3	13.660	4.304	-14.180	218
4	13.792	4.376	-14.276	265
5	14.000	4.544	-14.236	312
6	14.200	4.732	-14.192	359
7	14.364	4.836	-14.256	421
8	14.652	4.996	-14.272	468
9	14.860	5.192	-14.240	515
10	15.116	5.364	-14.272	562
11	15.444	5.568	-14.200	625
12	15.684	5.668	-14.304	671
13	15.960	5.748	-14.412	718
14	16.340	5.888	-14.332	765
15	16.616	6.104	-14.280	828
...	...	...	...	...
127	14.212	5.004	-13.572	6.515

O ambiente de *hardware* utilizado durante a coleta de dados para os experimentos pode ser visualizado na Fig. 16. A luva instrumentada e o rastreador de movimento são conectados, juntamente com uma *webcam*, a um laptop com a seguinte configuração: processador AMD Turion™X2 Dual Core (2,0 GHz), 2Gb RAM, executando Microsoft Windows® XP Professional SP3.

6.2 Especificação de software

Os componentes de *software* utilizados no ambiente experimental dividem-se em dois grupos, de acordo com suas funções: aquisição de dados (objetivando registrar a execução dos gestos pelos usuários) e reconhecimento/avaliação (visando classificar os gestos executados).

Figura 16: Configuração de hardware do ambiente experimental.

6.2.1 Aquisição de dados

Os dados utilizados na comparação dos métodos de reconhecimento foram obtidos a partir de um conjunto de usuários. Para evitar que esses usuários se aprofundassem em questões de configuração e funcionamento dos dispositivos instrumentados, e se concentrassem na geração dos dados necessários às comparações, foi desenvolvida uma aplicação com a finalidade específica de se comunicar com os dispositivos, coletar dados e organizá-los para avaliação posterior.

A aplicação foi desenvolvida de forma a facilitar a execução dos gestos pelos usuários. O objetivo maior foi evitar que os mesmos tivessem que aprender um vocabulário inteiro, relacionando seus gestos

componentes com uma descrição ou uma ação específica. A alternativa adotada baseou-se na idéia da *exemplificação*, com os usuários utilizando exemplos dos gestos a serem executados, sem ter que decorá-los.

Para que a exemplificação funcionasse adequadamente, a tela principal da aplicação de coleta de dados foi dividida em duas partes: na primeira (lado esquerdo), um componente capaz de executar *streams* de vídeo foi posicionado; nesse componente, os vídeos de exemplo a serem seguidos são executados, tanto para posturas quanto para trajetórias; na segunda (lado direito), o *stream* proveniente da *webcam* é exibido, oferecendo aos usuários um *feedback* visual imediato, de forma a auxiliá-los em eventuais correções das posturas e trajetórias executadas. Os demais controles disponibilizados pela aplicação foram dispostos em uma barra de menus e em uma barra de botões, respectivamente posicionadas nas partes superior e inferior da tela principal. Exemplos da aplicação em execução podem ser vistos na Fig. 17.

O processo de coleta de dados é iniciado com a calibragem da luva instrumentada, visando ajustar os sensores à anatomia da mão de cada usuário. A calibragem consiste na execução de uma sequência de posturas, que são apresentadas aos usuários na forma de um vídeo, por aproximadamente um minuto. Durante a execução das posturas, o *driver* da *data glove* armazena os valores máximos e mínimos de flexão atingidos pelo usuário, valores esses que serão utilizados posteriormente para o processamento dos dados do tipo *scaled*.

A mesma sequência de passos é executada para a aquisição de dados de posturas e trajetórias. No início de cada iteração de coleta, a lista de gestos é embaralhada, de forma a evitar que um mesmo gesto seja executado repetidamente. Em seguida, os vídeos com as posturas e trajetórias são exibidos, um a um, sendo imitados pelos usuários. É permitido aos usuários pausar e assistir ao vídeo da vez repetidamente, se necessário. O tipo de gesto que está sendo exibido guia o comportamento da aplicação: para posturas, um clique no botão *Record* captura os dados *raw* e *scaled* da luva instrumentada; para trajetórias, um primeiro clique no botão *Record* inicia a captura dos dados usando os *frames* de dados provenientes do rastreador de movimento, enquanto que um segundo clique no mesmo botão interrompe a captura. Os dados coletados são armazenados em arquivos texto individuais, identificados pelo número do usuário, tipo de dado (*raw* ou *scaled* para posturas e *fob* para trajetórias), número da postura ou trajetória e número da iteração. Dada a disponibilidade dos dados adquiridos pela *webcam*, *screenshots*

das posturas e *streams* de vídeo das trajetórias executadas são gravados juntamente com os dados instrumentados.

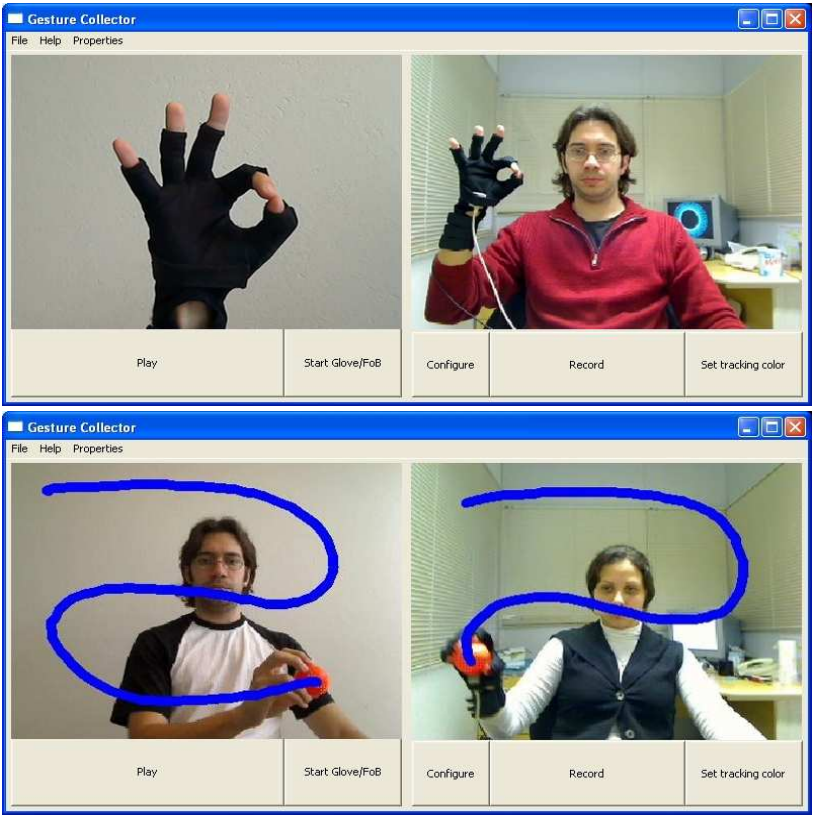

Figura 17: Aplicação para coleta de dados – postura e trajetória.

Para garantir o *feedback* visual aos usuários, a *webcam* é ligada no momento em que a aplicação inicia sua execução, e permanece ativa enquanto a mesma for utilizada. As posturas executadas podem ser visualizadas como *streams* de vídeo normais, e as trajetórias como *linhas de rastreamento*, geradas pela aplicação com base na identificação de uma cor específica. Durante a execução de uma trajetória, a aplicação traça uma linha de acordo com a movimentação do objeto que apresenta a cor a ser rastreada, permitindo sua visualização. Essa estratégia visa auxiliar os usuários, provendo uma base visual de

comparação entre a trajetória original (exibida no vídeo de modelo) e a trajetória que está sendo executada.

6.2.2 Reconhecimento/avaliação

Para a classificação dos dados coletados, foi escolhido um conjunto de técnicas para cada tipo de gesto. A escolha se deu com base nos métodos utilizados pelos trabalhos relacionados no capítulo cinco, adotando os critérios de relevância e acurácia obtida nos experimentos realizados.

As posturas foram avaliadas utilizando-se técnicas baseadas em redes neurais (implementadas com a biblioteca *Fast Artificial Neural Network* – FANN[14]), SVM (implementada com a biblioteca LIBSVM[15]) e em uma *KD-Tree*, computando resultados através de *nearest neighbor* com base na biblioteca *Approximate Nearest Neighbor* (ANN[16]); as trajetórias, por sua vez, foram avaliadas por implementações de redes neurais (FANN), HMM (implementado com a biblioteca *Accelerometer Gesture Recognizer* – AGR[17]) e uma heurística própria, definida especificamente para esta dissertação[18].

A heurística proposta neste trabalho baseia-se no mapeamento entre uma trajetória e um AFD – Autômato Finito Determinístico. O mapeamento transforma as menores unidades representativas de uma trajetória (os pontos no espaço tridimensional) nas menores unidades representativas de um AFD (os estados do autômato), resultando em uma estrutura semelhante à da Fig. 18. Assim, o autômato mantém a definição formal dada por HOPCROFT et al. (2001), sendo formado pela quíntupla $(Q, \Sigma, \delta, q_0, F)$:

- Q – conjunto finito de estados do autômato, mapeados de um-para-um com os pontos no espaço tridimensional da trajetória original;
- Σ – um conjunto de símbolos de entrada, formado por pontos no espaço tridimensional;
- δ – uma função de transição, que avalia o estado atual do autômato e o símbolo de entrada, gerando ou não uma mudança de estado de acordo com a avaliação;

[14] http://leenissen.dk/fann/
[15] http://www.csie.ntu.edu.tw/~cjlin/libsvm/
[16] http://www.cs.umd.edu/~mount/ANN/
[17] http://sourceforge.net/projects/agr/
[18] Os métodos selecionados foram implementados utilizando Microsoft Visual C++ 2005.

- q_0 – um estado inicial, correspondente ao primeiro ponto no espaço tridimensional da trajetória;
- F – um conjunto de estados finais, subconjunto de Q; no mapeamento executado, o estado final corresponde ao último ponto no espaço tridimensional da trajetória.

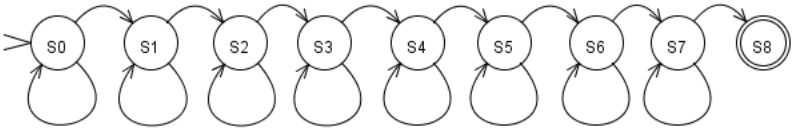

Figura 18: AFD criado a partir de uma trajetória.

Cada trajetória de treinamento (ou modelo) gera um autômato, compondo um conjunto de padrões de avaliação. No momento do reconhecimento, cada ponto no espaço tridimensional pertencente à trajetória executada é utilizado como símbolo de entrada, submetido à função de transição dos autômatos criados previamente. Para todos os estados dos autômatos (com exceção aos estados finais), há duas transições possíveis: uma que leva ao próximo estado, e outra que mantém o estado atual. Uma transição executada para o próximo estado indica que o símbolo de entrada avaliado (o ponto no espaço tridimensional) pertence ao padrão representado pelo autômato; uma transição para o mesmo estado indica que o símbolo de entrada não pertence a esse padrão. Após a avaliação do último símbolo de entrada disponível, o gesto reconhecido é aquele relacionado ao primeiro autômato que atingiu seu estado final, ou aquele relacionado ao autômato cujo estado atual é o mais próximo ao estado final.

A função de transição escolhida para a implementação da heurística proposta baseia-se em coordenadas esféricas. Essas coordenadas permitem localizar pontos em um espaço tridimensional com base em dois ângulos de orientação e uma medida de distância do ponto considerado ao ponto de origem do sistema de coordenadas. Os ângulos φ (*Fi*), θ (*Teta*) e a medida de distância r são calculados utilizando-se as coordenadas X, Y e Z do ponto a ser localizado, de acordo com as equações (1), (2) e (3).

$$\varphi = \arctan\left(\frac{\sqrt{X^2 + Y^2}}{Z}\right) \quad (1)$$

$$\theta = \arctan\left(\frac{Y}{X}\right) \quad (2)$$

$$r = \sqrt{X^2 + Y^2 + Z^2} \quad (3)$$

Para a heurística proposta, a medida de distância é ignorada. Apenas os ângulos de orientação são utilizados, de forma a identificar se a trajetória executada apresenta uma tendência a ter a mesma orientação espacial do padrão de avaliação. Os ângulos são calculados para cada símbolo de entrada e comparados com os ângulos previamente armazenados nos estados atuais dos autômatos. Caso os valores se encontrem dentro de um intervalo de aceitação, a orientação é considerada similar e os autômatos correspondentes têm seus estados atuais atualizados.

6.3 Vocabulário de gestos

Os gestos escolhidos para compor o vocabulário utilizado na presente dissertação foram selecionados a partir de diferentes trabalhos relacionados a interfaces gestuais, sendo adaptados quando necessário às características do *hardware* disponível. As posturas podem ser visualizadas na Fig. 19, e as trajetórias na Fig. 20.

Para as posturas, as características principais a serem observadas durante o reconhecimento são as flexões executadas pelos dedos, individualmente. As disposições de um dedo em relação aos outros dedos e as rotações e inclinações da mão em relação ao braço e à câmera não são capturadas pela *data glove* e, portanto, não têm relevância. As trajetórias, por sua vez, são executadas em um espaço tridimensional. Apesar do traçado das mesmas ser claramente bidimensional, diferenças relacionadas ao eixo representativo de profundidade são relevantes na computação dos resultados.

Com o ambiente experimental definido e organizado, executou-se a coleta de dados e iniciou-se o trabalho de comparação propriamente dito. Os resultados obtidos pela utilização da configuração de *hardware*

e *software*, bem como os resultados decorrentes da comparação entre métodos de reconhecimento, são apresentados no capítulo a seguir.

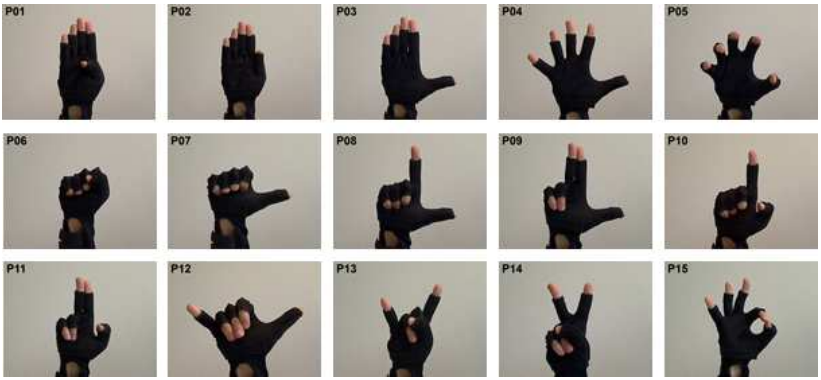

Figura 19: Posturas selecionadas para o vocabulário de gestos.

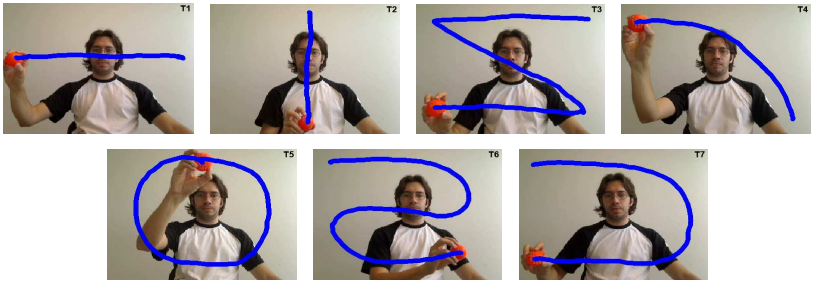

Figura 20: Trajetórias selecionadas para o vocabulário de gestos.

7. Resultados

O ambiente experimental descrito no capítulo anterior foi utilizado por um conjunto de usuários voluntários, responsáveis pela execução de posturas e trajetórias – base de treinamento e avaliação para a comparação entre as técnicas de reconhecimento. Os resultados da avaliação dos dados coletados são apresentados neste capítulo.

7.1 Organização e visualização dos dados coletados

O conjunto de dados coletado para o presente trabalho foi obtido a partir de 33 usuários voluntários, sendo 29 homens e 4 mulheres, com idades entre 19 e 50 anos, sem experiências anteriores com interfaces gestuais. Juntos, esses usuários contribuíram com 2.475 posturas (33 usuários executando 15 posturas em cinco iterações) e 1.155 trajetórias (33 usuários executando sete trajetórias em cinco iterações). Dadas as características dos dados instrumentados, é possível representá-los graficamente para a visualização da sua distribuição.

A Fig. 21 exibe os dados do tipo *raw* separados por postura. Os gráficos cruzam os valores obtidos com os sensores, respectivamente dispondo-os nos eixos vertical e horizontal. Uma análise visual dos gráficos permite observar a grande variância na distribuição dos dados em um mesmo sensor, identificável pelo intervalo de valores registrado. Isso se deve à falta de normalização dos dados obtidos, fato que diferencia consideravelmente posturas semelhantes, e que é justificado pelas diferenças anatômicas entre as mãos dos usuários.

A Fig. 22, por sua vez, exibe os dados do tipo *scaled* para as mesmas posturas. Visualmente, não parece haver diferença: a variância existe, e se assemelha àquela observada na Fig. 21; há, porém, um detalhe que deve ser levado em consideração: a significativa diferença de escala presente no eixo dos valores obtidos. Os diferentes tipos de dados seguem escalas específicas, cujos valores correspondem aos intervalos citados no capítulo anterior. A primeira impressão decorrente de uma avaliação visual é de que os dados do tipo *scaled*, por estarem distribuídos em um intervalo mais limitado, são mais semelhantes; como conseqüência, cada postura teria uma identidade melhor definida e, por fim, mais facilmente classificável em relação às demais.

As trajetórias coletadas durante a aquisição de dados podem ser visualizadas na Fig. 23. Nesta figura, apesar do grande volume de linhas de rastreamento presentes, é possível identificar a tendência das

trajetórias e relacioná-las aos modelos visualizados na Fig. 20. É importante ressaltar que os gráficos da Fig. 23 apresentam os dados originalmente coletados, sem filtros ou normalizações, dispostos nos eixos X e Y – o eixo Z foi suprimido para facilitar a visualização dos padrões.

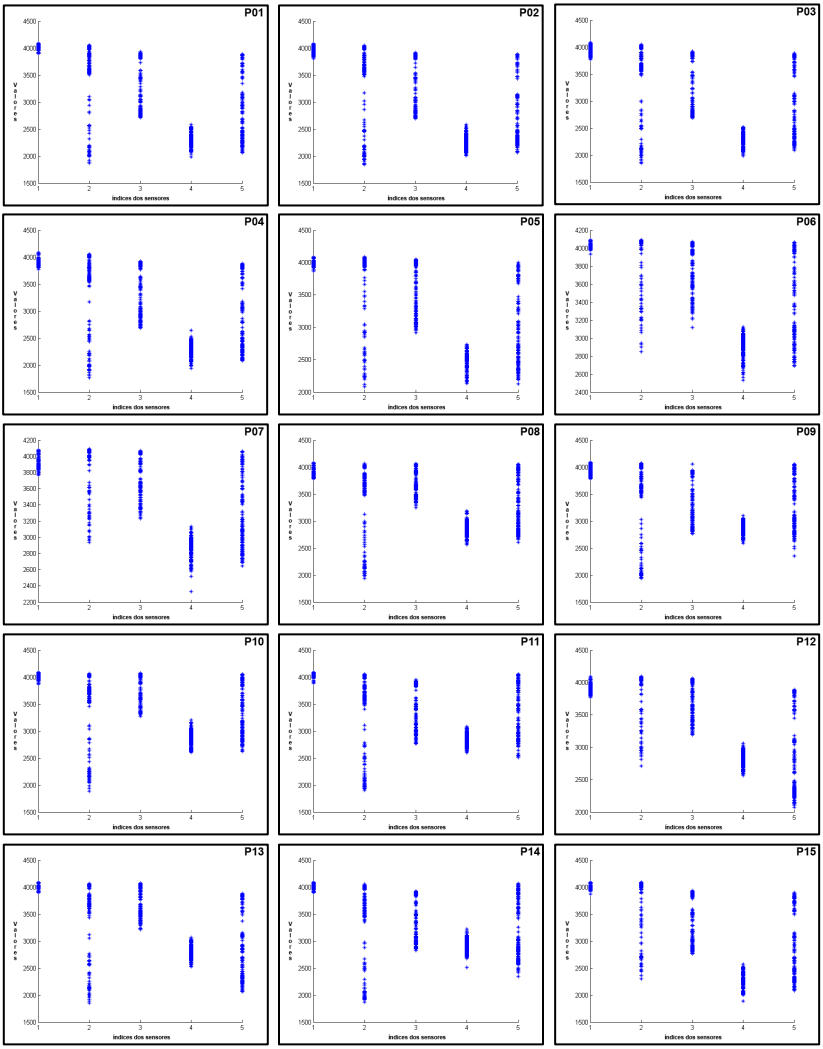

Figura 21: Representação gráfica dos dados do tipo *raw*.

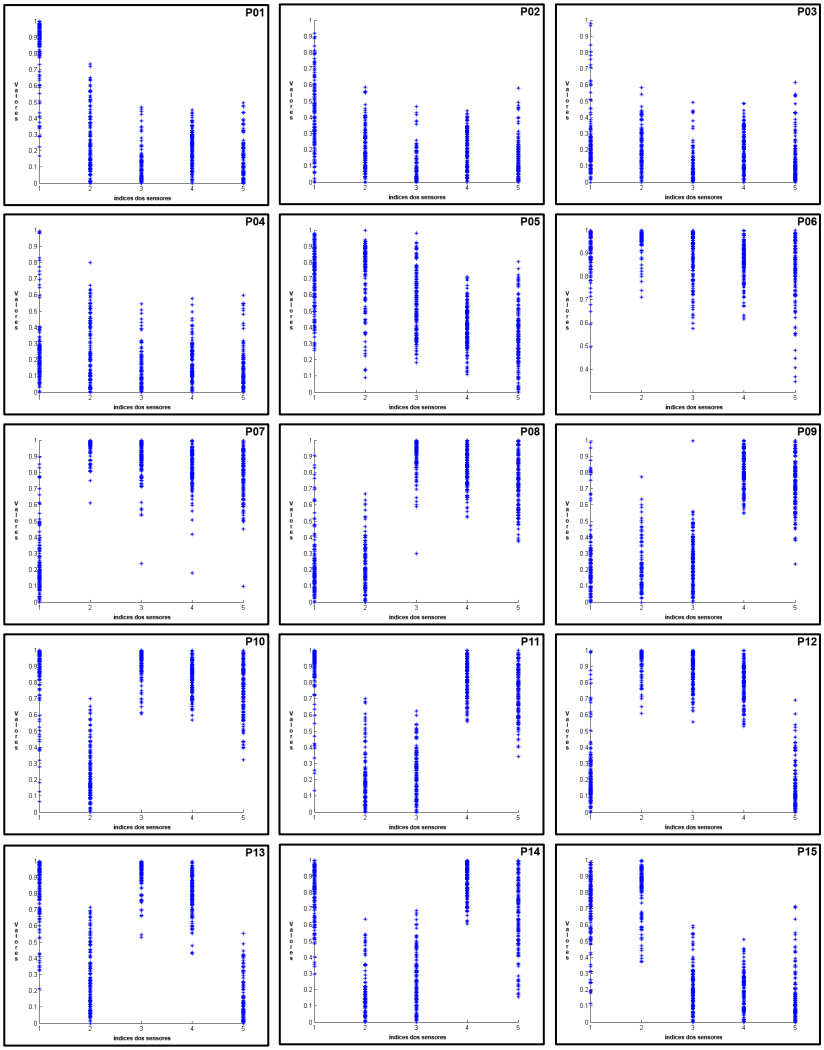

Figura 22: Representação gráfica dos dados do tipo *scaled*.

É possível, assim, identificar duas características inerentes à execução das trajetórias. A primeira corresponde à forma como os gestos são executados: apesar de existir um modelo com uma trajetória bem definida e uma velocidade observável, cada usuário interpreta essas duas características de forma particular e as mimetiza também de forma

particular, o que gera uma grande diversidade de padrões. A segunda diz respeito à inicialização e ao término dos gestos. Nos gráficos da Fig. 23, é possível identificar pontos no espaço tridimensional que não obedecem aos traçados originais; boa parte desses pontos coincide com as posições inicial e final das trajetórias, o que indica um desvio de atenção por parte dos usuários gerado pela necessidade de iniciar e terminar o gesto. A existência desses pontos pode, por consequência, interferir nos resultados dos reconhecimentos – pelo fato de gerarem variações que diminuem a similaridade da trajetória executada em relação aos padrões utilizados na classificação.

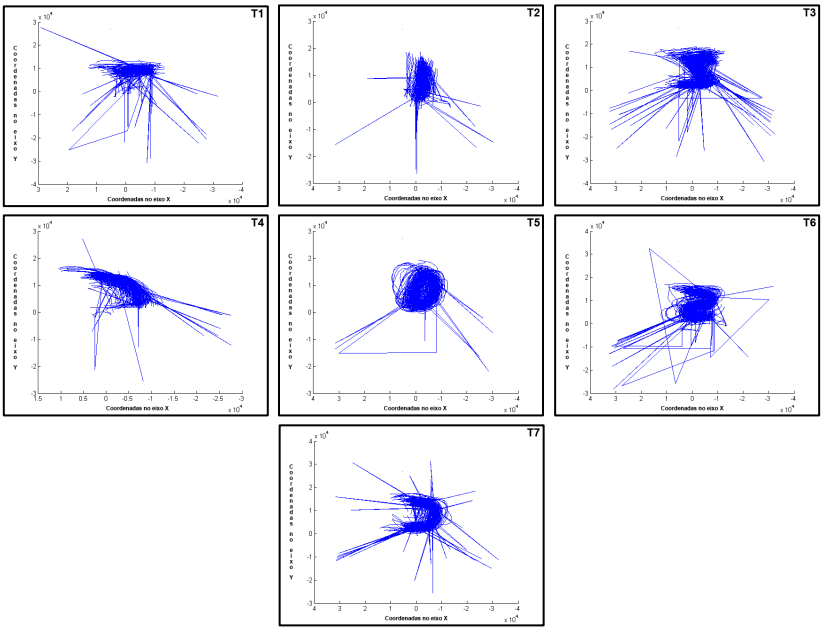

Figura 23: Representação gráfica das trajetórias.

7.2 Reconhecimento e classificação dos gestos

Os métodos de reconhecimento escolhidos para ambos os tipos de gestos foram submetidos a um processo derivado da técnica conhecida como *random subsampling validation*, na qual o conjunto de dados original é particionado em dois subconjuntos: um para treinamento, com

1/3 dos dados (aleatoriamente selecionados), e um para avaliação (formado pelos 2/3 restantes). A avaliação foi executada com 10 iterações, mantendo-se como resultado o número total de reconhecimentos. A definição do tamanho dos conjuntos de treinamento e avaliação, bem como a escolha da técnica de validação e o número de iterações aplicado, foi feita arbitrariamente.

Objetivando demonstrar a importância da escolha dos dados componentes do conjunto de treinamento, duas estratégias de particionamento foram adotadas neste trabalho. A primeira, chamada EP1 (estratégia de particionamento 1), seleciona como dados de treinamento 1/3 do conjunto total de gestos igualmente distribuídos entre suas classes, randomicamente. A segunda, chamada EP2 (estratégia de particionamento 2), seleciona como dados de treinamento todos os gestos executados por 1/3 dos usuários, selecionados aleatoriamente. Como resultado da aplicação das estratégias citadas, espera-se identificar a melhor abordagem para que se obtenha uma melhor generalização, observável diretamente na acurácia das técnicas escolhidas.

7.2.1 Posturas

O método baseado em *KD-Tree* e *nearest neighbor* foi avaliado em primeiro lugar, tendo seus resultados organizados na forma de matrizes de confusão (Quadros 4 e 5), respectivamente para as estratégias EP1 e EP2. As posturas esperadas são dispostas em linhas, enquanto que as posturas classificadas (os resultados obtidos) são dispostas em colunas; essa disposição permite relacionar a diagonal principal das matrizes (em destaque) com os verdadeiros positivos, deixando as demais células com os falsos positivos. Cada célula exibe dois valores: o primeiro corresponde aos resultados obtidos com os dados do tipo *raw*, e o segundo aos resultados obtidos com os dados do tipo *scaled*. A precisão do método pode ser visualizada na última coluna da matriz – individualmente por postura, ou coletivamente para todo o conjunto.

A avaliação do método baseado em rede neural foi executada na sequência, e seus resultados podem ser visualizados nos Quadros 6 e 7. A rede utilizada foi estruturada com cinco camadas: uma camada de entrada com cinco nodos, três camadas ocultas com 18 nodos em cada camada e uma camada de saída com 15 nodos. O treinamento foi executado com 1.000 épocas, utilizando o algoritmo *backpropagation* com uma taxa de aprendizagem de 0,5. Em comparação à técnica

baseada em *nearest neighbor*, a rede neural apresentou uma precisão levemente inferior na avaliação dos dados do tipo *scaled* (uma diferença de 6,4% e 0,55%, respectivamente, para as estratégias EP1 e EP2). A maior diferença percebida, porém, reside na classificação dos dados do tipo *raw*: enquanto a técnica baseada em *nearest neighbor* atingiu uma acurácia máxima de 70,49%, a rede neural obteve apenas 17,98%. Uma explicação possível para essa diferença está na escala dos valores *raw*: de acordo com SARLE (1997), redes neurais e outras estruturas de classificação podem apresentar problemas de natureza numérica quando submetidas a valores situados em intervalos esparsos.

Quadro 4: Método baseado em *nearest neighbor*/particionamento EP1.

		Classificado (*raw/scaled*)															Precisão (%)
		P01	**P02**	**P03**	**P04**	**P05**	**P06**	**P07**	**P08**	**P09**	**P10**	**P11**	**P12**	**P13**	**P14**	**P15**	
Esperado (*raw/scaled*)	**P01**	**770/828**	123/132	49/30	58/33	41/18	0/0	1/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	58/59	70,00/75,27
	P02	145/142	**446/553**	273/247	204/135	18/11	0/0	0/2	0/0	1/2	0/0	0/0	0/0	0/0	1/0	12/8	40,55/50,27
	P03	95/39	263/222	**469/518**	248/302	14/5	0/0	0/0	0/0	1/6	0/0	0/0	0/0	0/0	1/0	9/8	42,64/47,09
	P04	58/47	206/140	247/315	**526/548**	24/9	0/0	0/0	4/4	3/6	4/4	0/0	0/0	0/0	0/0	28/27	47,82/49,82
	P05	40/27	16/28	5/7	21/7	**873/932**	17/10	5/0	6/0	2/0	2/8	4/0	0/5	13/0	0/0	96/76	79,36/84,73
	P06	1/0	0/0	0/0	0/0	14/21	**905/1053**	57/25	0/0	4/0	42/1	14/0	30/0	8/0	25/0	0/0	82,27/95,73
	P07	0/1	0/4	0/0	0/2	14/2	75/40	**900/1035**	34/0	23/0	23/0	2/0	24/16	4/0	1/0	0/0	81,82/94,09
	P08	0/0	0/0	0/0	1/0	3/0	6/0	52/1	**861/973**	20/16	130/98	6/0	10/0	10/12	1/0	0/0	78,27/88,45
	P09	0/0	0/0	0/0	1/0	4/1	2/0	15/2	10/7	**817/958**	4/0	128/77	1/0	0/0	118/55	0/0	74,27/87,09
	P10	0/0	0/0	0/0	0/1	1/13	64/1	10/0	120/90	2/2	**836/964**	9/3	0/0	52/19	6/7	0/0	76,00/87,64
	P11	0/0	1/0	0/0	0/0	4/1	13/0	0/0	0/0	81/51	3/0	**678/720**	0/0	0/0	320/328	0/0	61,64/65,45
	P12	0/0	0/0	0/0	0/0	8/0	14/10	35/25	0/0	0/0	8/0	0/0	**998/1064**	37/1	0/0	0/0	90,73/96,73
	P13	0/0	0/0	0/2	0/3	12/2	2/0	2/1	5/14	0/0	59/17	0/0	36/3	**984/1058**	0/0	0/0	89,45/96,18
	P14	0/0	0/0	5/0	2/0	1/5	11/0	0/0	0/0	70/34	16/2	278/336	0/0	0/0	**717/723**	0/0	65,18/65,73
	P15	77/58	24/7	10/2	56/11	80/85	0/0	2/0	0/0	0/2	0/0	0/0	0/0	0/0	0/0	**851/935**	77,36/85,00
																Precisão média (%)	**70,49/77,95**

Quadro 5: Método baseado em *nearest neighbor*/particionamento EP2.

		Classificado (*raw/scaled*)															Precisão (%)
		P01	**P02**	**P03**	**P04**	**P05**	**P06**	**P07**	**P08**	**P09**	**P10**	**P11**	**P12**	**P13**	**P14**	**P15**	
Esperado (*raw/scaled*)	**P01**	**405/784**	152/171	87/28	104/36	123/32	1/0	1/1	0/0	0/0	0/0	2/0	1/0	6/0	1/0	217/48	36,82/71,27
	P02	154/169	**228/454**	215/240	292/198	94/15	0/0	2/4	0/0	0/4	0/0	1/0	0/0	2/6	1/0	111/10	20,73/41,27
	P03	73/69	185/234	**301/428**	330/345	89/6	1/0	4/0	0/0	0/9	0/0	0/2	1/0	3/3	1/0	112/4	27,36/38,91
	P04	40/61	186/166	241/340	**393/472**	104/16	0/0	1/0	6/5	2/15	0/0	0/3	0/0	0/0	0/0	127/22	35,73/42,91
	P05	79/68	28/32	26/7	38/15	**537/812**	22/19	31/3	14/0	2/1	11/40	2/1	18/9	70/0	6/0	216/93	48,82/73,82
	P06	0/0	0/0	0/0	0/0	57/28	**597/996**	81/43	9/0	17/0	118/3	38/0	117/30	16/0	50/0	0/0	54,27/90,55
	P07	1/0	1/3	1/0	0/3	72/9	141/37	**580/1022**	105/0	45/0	25/1	0/0	117/25	2/0	10/0	0/0	52,73/92,91
	P08	0/0	0/0	0/0	3/0	31/1	30/0	93/9	**654/958**	36/11	123/110	2/0	7/0	107/11	14/0	0/0	59,45/87,09
	P09	3/0	0/1	0/0	0/4	17/0	22/0	99/0	39/14	**592/986**	24/0	159/43	0/0	2/0	143/52	0/0	53,82/89,64
	P10	0/0	0/0	0/0	1/0	10/12	118/0	11/18	124/104	3/3	**641/918**	14/0	3/0	140/37	35/8	0/0	58,27/83,45
	P11	0/0	0/0	0/0	0/0	19/3	99/1	11/0	12/0	76/105	34/2	**527/580**	1/0	4/0	317/409	0/0	47,91/52,73
	P12	0/0	0/0	0/0	1/0	16/29	57/9	52/47	23/0	1/0	11/0	0/0	**802/1015**	136/0	1/0	0/0	72,91/92,27
	P13	2/1	0/3	0/7	1/6	16/5	18/0	3/2	18/47	0/0	122/43	0/0	107/4	**813/982**	0/0	0/0	73,91/89,27
	P14	0/0	2/0	1/0	2/0	3/3	75/0	6/0	11/0	81/74	54/3	372/424	2/0	7/0	**482/596**	2/0	43,82/54,18
	P15	123/124	100/10	35/3	60/27	254/134	0/0	4/0	0/0	0/5	0/0	0/0	0/0	0/0	1/0	**523/797**	47,55/72,45
																Precisão média (%)	**48,94/71,52**

Quadro 6: Método baseado em rede neural com particionamento EP1.

Esperado (*raw*/*scaled*)	Classificado (*raw*/*scaled*)															Precisão (%)
	P01	P02	P03	P04	P05	P06	P07	P08	P09	P10	P11	P12	P13	P14	P15	
P01	**239/891**	110/34	122/29	184/12	0/22	27/0	31/0	75/0	0/0	0/0	41/8	85/1	3/6	4/1	179/96	21,73/81,00
P02	229/327	**110/174**	123/318	178/156	1/21	27/0	40/0	74/0	0/36	1/0	47/3	87/1	4/11	3/0	176/53	10,00/15,82
P03	235/108	110/61	**120/561**	181/270	1/13	27/0	40/3	74/0	0/50	1/0	46/7	87/2	5/4	3/2	170/19	10,91/51,00
P04	217/59	109/35	124/471	**183/313**	0/48	27/0	43/6	71/8	0/43	1/0	40/0	86/10	4/2	3/0	192/105	16,64/28,45
P05	174/36	100/7	114/5	111/16	**0/694**	28/114	98/12	81/0	0/1	10/3	32/0	144/45	20/9	6/0	182/158	0,00/63,09
P06	71/0	45/0	110/0	13/0	1/7	**43/1074**	167/4	122/0	30/0	133/0	71/0	163/10	3/5	100/0	28/0	3,91/97,64
P07	31/0	38/0	107/0	13/0	0/1	38/79	**248/998**	117/0	29/0	108/0	55/0	190/15	6/0	76/0	44/7	22,55/90,73
P08	45/0	14/0	73/0	3/0	0/0	44/1	138/13	**255/975**	31/13	234/96	35/0	130/0	28/2	65/0	5/0	23,18/88,64
P09	13/1	27/0	103/0	8/0	0/0	47/8	79/5	96/13	**140/975**	77/1	199/71	84/0	3/0	224/26	0/0	12,73/88,64
P10	78/0	22/0	82/0	6/0	1/3	42/18	123/23	218/59	21/3	**239/946**	37/2	128/1	20/45	73/0	10/0	21,73/86,00
P11	18/1	34/0	107/0	17/0	0/6	46/8	68/0	96/0	118/61	71/25	**215/701**	96/0	3/0	208/298	3/0	19,55/63,73
P12	45/0	12/0	104/0	1/0	1/8	26/10	123/27	113/0	0/0	60/0	4/0	**502/1046**	49/9	12/0	48/0	45,64/95,09
P13	72/0	15/0	69/0	1/0	0/0	26/0	84/0	164/33	0/0	132/33	14/0	265/8	**183/1026**	17/0	58/0	16,64/93,27
P14	23/7	30/0	101/0	14/0	0/3	47/4	62/1	106/3	133/106	80/58	196/451	90/0	8/6	**208/461**	2/0	18,91/41,91
P15	184/63	110/0	114/2	200/6	0/54	27/3	38/1	34/0	0/0	0/1	13/0	99/0	0/0	0/0	**281/970**	25,55/88,18
															Precisão média (%)	**17,98/71,55**

Quadro 7: Método baseado em rede neural com particionamento EP2.

Esperado (*raw*/*scaled*)	Classificado (*raw*/*scaled*)															Precisão (%)
	P01	P02	P03	P04	P05	P06	P07	P08	P09	P10	P11	P12	P13	P14	P15	
P01	**93/910**	0/29	3/28	196/21	0/33	39/2	1/0	0/0	68/1	22/4	116/4	46/0	218/0	149/1	149/67	8,45/82,73
P02	82/322	**0/189**	0/307	197/146	0/54	36/0	6/0	0/0	68/32	15/7	123/1	47/0	225/6	149/0	152/36	0,00/17,18
P03	81/115	0/62	**1/592**	193/222	0/26	36/0	4/0	2/2	70/50	11/8	123/7	44/0	226/7	151/2	158/7	0,09/53,82
P04	76/55	1/34	0/520	**208/291**	0/75	31/0	10/3	1/5	68/41	8/10	104/0	42/2	221/2	141/0	189/62	18,91/26,45
P05	29/78	0/1	0/11	162/15	**0/667**	75/101	23/4	0/0	65/0	20/17	119/1	61/35	242/21	155/0	149/149	0,00/60,64
P06	0/0	0/0	0/0	6/0	0/5	**158/1080**	13/2	21/0	115/0	33/9	223/0	48/3	214/1	264/0	5/0	14,36/98,18
P07	0/0	0/0	0/0	9/1	0/3	157/94	**49/982**	39/1	109/0	22/0	177/0	49/16	218/0	252/0	19/3	4,45/89,27
P08	0/0	1/0	0/0	19/0	0/0	138/1	2/8	**91/968**	118/15	50/98	167/1	27/0	267/8	216/1	4/0	8,27/88,00
P09	0/1	0/0	0/3	0/0	0/0	103/8	0/1	15/9	**180/967**	33/2	255/64	26/0	168/0	315/45	5/0	16,36/87,91
P10	1/0	1/0	0/0	15/0	0/4	147/15	0/10	59/56	109/4	**56/981**	176/2	17/0	276/25	242/3	1/0	5,09/89,18
P11	3/5	0/0	0/0	0/0	0/2	107/18	0/0	5/0	188/60	34/25	**258/653**	13/0	158/0	329/337	5/0	23,45/59,36
P12	0/0	0/0	0/0	60/0	0/18	187/7	12/37	4/0	29/0	5/3	53/0	**172/1031**	394/4	107/0	77/0	15,64/93,73
P13	12/0	0/0	0/0	88/0	0/3	158/2	2/0	12/27	39/0	27/36	60/0	61/5	**492/1027**	136/0	13/0	44,73/93,36
P14	1/17	0/0	0/3	0/0	0/1	109/5	1/0	4/1	162/116	28/45	252/427	22/0	196/2	**324/483**	1/0	29,45/43,91
P15	39/89	0/0	3/4	200/5	0/95	12/4	13/6	0/0	41/0	15/0	107/4	88/4	119/0	123/0	**340/889**	30,91/80,82
															Precisão média (%)	**14,68/70,97**

Os Quadros 8 e 9 exibem os dados obtidos pelo método baseado em SVM. O classificador construído utilizou um *kernel* do tipo RBF (*Radial Basis Function*), atingindo os melhores resultados na avaliação dos dados do tipo *scaled* (79,13%), e os piores resultados na avaliação dos dados do tipo *raw* (6,81%). Dada a sua organização, a SVM é sujeita a restrições semelhantes àquelas apresentadas pela rede neural, no tocante aos intervalos de valores utilizados.

Uma análise das matrizes de confusão geradas durante o processo de avaliação permite algumas constatações. São elas:

- *tipos de dados* – os melhores resultados foram obtidos com os dados do tipo *scaled*. Os dois fatores que contribuem para isso são a calibragem da *data glove* (que permite uma aquisição de dados otimizada de acordo com a anatomia das mãos dos

usuários) e a diferença de escala em relação aos dados *raw*, diferença essa relevante para a rede neural e a SVM;

- *estratégias de particionamento* – a adoção da estratégia EP1 resultou em melhores resultados quando compara à estratégia EP2. Isso ocorre pela maior heterogeneidade obtida quando instâncias de gestos geradas por um número variável de usuários são escolhidas; a limitação do número de usuários (característica principal da estratégia EP2) acaba por influenciar na generalização dos classificadores.

Quadro 8: Método baseado em SVM com particionamento EP1.

		Classificado (*raw*/*scaled*)															Precisão (%)
		P01	**P02**	**P03**	**P04**	**P05**	**P06**	**P07**	**P08**	**P09**	**P10**	**P11**	**P12**	**P13**	**P14**	**P15**	
Esperado (*raw*/*scaled*)	**P01**	**11/884**	0/114	0/10	4/28	0/23	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	1085/41	1,00/80,36
	P02	0/89	**0/605**	5/226	0/156	0/20	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	1095/4	0,00/55,00
	P03	0/42	4/179	**6/549**	0/326	0/4	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	1090/0	0,55/49,91
	P04	2/43	0/124	0/363	**3/520**	0/16	0/0	0/0	0/4	0/2	0/1	0/1	0/0	0/2	0/0	1095/24	0,27/47,27
	P05	0/26	0/27	0/8	0/6	**12/934**	0/8	0/0	0/0	0/0	0/9	0/0	0/5	0/2	0/0	1088/75	1,09/84,91
	P06	0/0	0/0	0/0	0/0	0/21	**43/1053**	2/18	0/0	0/0	0/1	0/0	0/7	0/0	0/0	1055/0	3,91/95,73
	P07	0/0	0/0	0/0	0/6	0/0	2/42	**20/1028**	0/0	0/0	0/3	0/0	0/21	0/0	0/0	1078/0	1,82/93,45
	P08	0/0	0/0	0/0	0/0	0/0	0/0	0/5	**19/1008**	0/6	0/71	0/0	0/0	0/10	0/0	1081/0	1,73/91,64
	P09	0/0	0/0	0/0	0/0	0/1	0/0	0/1	0/6	**3/979**	0/0	0/64	0/0	0/0	0/49	1097/0	0,27/89,00
	P10	0/0	0/0	0/0	0/0	0/5	0/0	0/6	0/89	0/0	**13/969**	0/0	0/0	1/29	0/2	1086/0	1,18/88,09
	P11	0/0	0/0	0/0	0/0	0/4	0/0	0/0	0/0	0/73	0/0	**10/786**	0/0	0/0	2/237	1088/0	0,91/71,45
	P12	0/0	0/0	0/0	0/0	0/0	0/14	0/20	0/0	0/0	0/0	0/0	**14/1066**	0/0	0/0	1086/0	1,27/96,91
	P13	0/0	0/0	0/3	0/2	0/0	0/1	0/0	0/7	0/0	4/18	0/0	0/4	**21/1065**	0/0	1075/0	1,91/96,82
	P14	0/0	0/0	0/0	0/0	0/6	0/0	0/0	0/2	5/27	0/3	2/414	0/0	0/0	**6/648**	1087/0	0,55/58,91
	P15	0/84	0/3	0/0	0/12	0/39	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	**1100/962**	100,00/87,45
																Precisão média (%)	**7,76/79,13**

Quadro 9: Método baseado em SVM com particionamento EP2.

		Classificado (*raw*/*scaled*)															Precisão (%)
		P01	**P02**	**P03**	**P04**	**P05**	**P06**	**P07**	**P08**	**P09**	**P10**	**P11**	**P12**	**P13**	**P14**	**P15**	
Esperado (*raw*/*scaled*)	**P01**	**0/857**	0/92	0/22	1/42	0/34	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	1099/53	0,00/77,91
	P02	0/145	**0/528**	2/218	0/179	0/25	0/0	0/0	0/0	0/1	0/0	0/0	0/0	0/0	0/0	1098/4	0,00/48,00
	P03	0/67	3/208	**5/493**	0/320	0/7	0/0	0/0	0/0	0/3	0/0	0/0	0/0	0/1	0/0	1092/1	0,45/44,82
	P04	4/63	0/163	0/383	**0/431**	0/21	0/0	0/0	0/7	0/12	0/0	0/0	0/0	0/0	0/0	1096/20	0,00/39,18
	P05	0/68	0/30	0/5	0/27	**0/835**	0/21	0/0	0/1	0/3	0/31	0/1	0/8	0/4	0/1	1100/65	0,00/75,91
	P06	0/0	0/0	0/0	0/0	0/32	**5/1023**	3/13	0/0	0/0	0/6	0/0	0/24	0/2	0/0	1092/0	0,45/93,00
	P07	0/0	0/1	0/0	0/6	0/17	2/69	**0/963**	0/2	0/2	0/4	0/0	0/36	0/0	0/0	1098/0	0,00/87,55
	P08	0/0	0/0	0/0	0/0	0/0	0/0	0/7	**0/967**	0/14	0/97	0/0	0/0	0/15	0/0	1100/0	0,00/87,91
	P09	0/0	0/0	0/0	0/2	0/1	0/0	0/1	0/6	**0/953**	0/0	0/80	0/0	0/0	2/57	1098/0	0,00/86,64
	P10	0/0	0/0	0/0	0/0	0/4	0/0	0/11	0/78	0/0	**5/946**	0/2	0/0	0/46	0/13	1095/0	0,45/86,00
	P11	0/0	0/0	0/0	0/0	0/5	0/0	0/0	0/0	0/55	0/1	**0/767**	0/0	0/0	0/272	1100/0	0,00/69,73
	P12	0/0	0/0	0/0	0/0	0/16	0/9	0/24	0/0	0/0	0/0	0/0	**0/1045**	0/6	0/0	1100/0	0,00/95,00
	P13	0/2	0/0	0/4	0/6	0/6	0/0	0/2	0/29	0/0	5/50	0/0	0/4	**8/997**	0/0	1087/0	0,73/90,64
	P14	0/0	0/0	0/0	0/0	0/3	0/0	0/0	0/0	3/46	0/3	0/468	0/0	0/0	**0/580**	1097/0	0,00/52,73
	P15	0/124	0/10	0/1	0/30	0/102	0/2	0/0	0/0	0/4	0/0	0/0	0/0	0/0	0/0	**1100/827**	100,00/75,18
																Precisão média (%)	**6,81/74,01**

Uma terceira constatação diz respeito ao vocabulário de gestos. Algumas posturas apresentaram baixas taxas de reconhecimento; como exemplos, podem ser citadas as posturas rotuladas como P02, P03, P04, P11 e P14. Uma análise visual dessas posturas na Fig. 19 permite identificar semelhanças (em termos de flexões dos dedos) entre P02,

P03 e P04, bem como entre P11 e P14. Essas semelhanças visuais são observáveis também nos dados adquiridos, o que pode levar a falsos positivos. Os Quadros 10, 11 e 12 exibem os resultados obtidos quando as posturas semelhantes são agrupadas, utilizando a estratégia de particionamento mais efetiva (EP1) e o tipo de dado *scaled*. Foram observadas melhoras na precisão de todos os métodos: 11,8% para o método baseado em *nearest neighbor*, 14,34% para a rede neural e 11,83% para a SVM. Um resumo dos resultados obtidos pode ser visualizado na Fig. 24.

Quadro 10: Método baseado em *nearest neighbor* com gestos agrupados.

		Classificado (*scaled*)												Precisão (%)
		P01	**P02/P03/P04**	**P05**	**P06**	**P07**	**P08**	**P09**	**P10**	**P11/P14**	**P12**	**P13**	**P15**	
Esperado (*scaled*)	**P01**	**828**	195	18	0	0	0	0	0	0	0	0	59	75,27
	P02/P03/P04	228	**2980**	25	0	2	4	14	4	0	0	0	43	90,30
	P05	27	42	**932**	10	0	0	0	8	0	5	0	76	84,73
	P06	0	0	21	**1053**	25	0	0	1	0	0	0	0	95,73
	P07	1	6	2	40	**1035**	0	0	0	0	16	0	0	94,09
	P08	0	0	0	0	1	**973**	16	98	0	0	12	0	88,45
	P09	0	0	1	0	2	7	**958**	0	132	0	0	0	87,09
	P10	0	1	13	1	0	90	2	**964**	10	0	19	0	87,64
	P11/P14	0	0	6	0	0	0	85	2	**2107**	0	0	0	95,77
	P12	0	0	0	10	25	0	0	0	0	**1064**	1	0	96,73
	P13	0	5	2	0	1	14	0	17	0	3	**1058**	0	96,18
	P15	58	20	85	0	0	0	2	0	0	0	0	**935**	85,00
													Precisão média (%)	**89,75**

Quadro 11: Método baseado em rede neural com gestos agrupados.

		Classificado (*scaled*)												Precisão (%)
		P01	**P02/P03/P04**	**P05**	**P06**	**P07**	**P08**	**P09**	**P10**	**P11/P14**	**P12**	**P13**	**P15**	
Esperado (*scaled*)	**P01**	**891**	75	22	0	0	0	0	0	9	1	6	96	81,00
	P02/P03/P04	494	**2359**	82	0	9	8	129	0	12	13	17	177	71,48
	P05	36	28	**694**	114	12	0	1	3	0	45	9	158	63,09
	P06	0	0	7	**1074**	4	0	0	0	0	10	5	0	97,64
	P07	0	0	1	79	**998**	0	0	0	0	15	0	7	90,73
	P08	0	0	0	1	13	**975**	13	96	0	0	2	0	88,64
	P09	1	0	0	8	5	13	**975**	1	97	0	0	0	88,64
	P10	0	0	3	18	23	59	3	**946**	2	1	45	0	86,00
	P11/P14	8	0	9	12	1	3	167	83	**1911**	0	6	0	86,86
	P12	0	0	8	10	27	0	0	0	0	**1046**	9	0	95,09
	P13	0	0	0	0	0	33	0	33	0	8	**1026**	0	93,27
	P15	63	8	54	3	1	0	0	1	0	0	0	**970**	88,18
													Precisão média (%)	**85,89**

Quadro 12: Método baseado em SVM com gestos agrupados.

Esperado (*scaled*)	Classificado (*scaled*)												Precisão (%)
	P01	P02/P03/P04	P05	P06	P07	P08	P09	P10	P11/P14	P12	P13	P15	
P01	**884**	152	23	0	0	0	0	0	0	0	0	41	80,36
P02/P03/P04	174	**3048**	40	0	0	4	2	1	1	0	2	28	92,36
P05	26	41	**934**	8	0	0	0	9	0	5	2	75	84,91
P06	0	0	21	**1053**	18	0	0	1	0	7	0	0	95,73
P07	0	6	0	42	**1028**	0	0	3	0	21	0	0	93,45
P08	0	0	0	0	5	**1008**	6	71	0	0	10	0	91,64
P09	0	0	1	0	1	6	**979**	0	113	0	0	0	89,00
P10	0	0	5	0	6	89	0	**969**	2	0	29	0	88,09
P11/P14	0	0	10	0	0	2	100	3	**2085**	0	0	0	94,77
P12	0	0	0	14	20	0	0	0	0	**1066**	0	0	96,91
P13	0	5	0	1	0	7	0	18	0	4	**1065**	0	96,82
P15	84	15	39	0	0	0	0	0	0	0	0	**962**	87,45
												Precisão média (%)	**90,96**

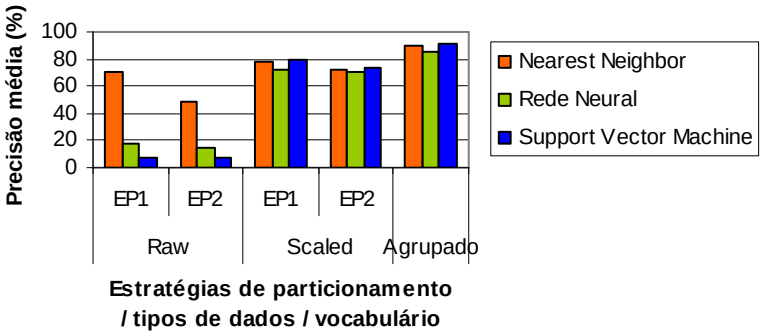

Figura 24: Resultado da avaliação das posturas.

Em relação ao desempenho, foram mensuradas as etapas de treinamento e reconhecimento. A etapa de treinamento compreende a carga dos padrões gestuais em memória, a construção das estruturas de dados responsáveis pelo armazenamento dos dados (de acordo com cada um dos classificadores) e a execução do algoritmo de treinamento. O método baseado em *nearest neighbor* não possui uma etapa de treinamento, visto que a *KD-Tree* utilizada se organiza em tempo de construção, não necessitando de adaptações posteriores. A etapa de reconhecimento totaliza o tempo necessário à carga do gesto a ser classificado em memória e o tempo gasto pelo classificador para gerar o resultado.

Os Quadros 13 e 14 apresentam os resultados obtidos, organizando-os em termos de *tempo médio de treinamento* (TMT) e *tempo médio de avaliação* (TMA), bem como o *desvio padrão* (DP) para todas as

técnicas. De acordo com os resultados, o uso dos tipos de dados *raw* e *scaled* gera tempos semelhantes para treinamento e reconhecimento. A exceção ocorre com a SVM, cujos tempos de treinamento e avaliação apresentam uma redução máxima de, respectivamente, 57,84% e 42,27% quando são utilizados dados do tipo *scaled*. A Fig. 25 exibe a redução para o TMA.

Quadro 13: Desempenho das etapas de treinamento e avaliação – dados *raw*.

	Desempenho (*raw*)			
	EP1		**EP2**	
	TMT/DP	**TMA/DP**	**TMT/DP**	**TMA/DP**
Nearest Neighbor	173,4ms/15,52ms	229,82µs/322,78µs	169,7ms/14,63ms	239,55µs/368,47µs
Rede Neural	40,78s/0,08s	264,34µs/220,20µs	41,6s/0,03s	259,69µs/188,4µs
Support Vector Machine	1,12s/0,08s	859,32µs/261,55µs	1,11s/0,07s	809,62µs/311,87µs

Quadro 14: Desempenho das etapas de treinamento e avaliação – dados *scaled*.

	Desempenho (*scaled*)			
	EP1		**EP2**	
	TMT/DP	**TMA/DP**	**TMT/DP**	**TMA/DP**
Nearest Neighbor	176,1ms/14ms	262,32µs/154,71µs	177,1ms/18,72ms	251,62µs/204,67µs
Rede Neural	40,68s/0,05s	272,19µs/154,65µs	41,59s/0,07s	268,08µs/188,48µs
Support Vector Machine	472,2ms/14,65ms	522,68µs/407,8µs	468,4ms/20,52ms	467,37µs/409,16µs

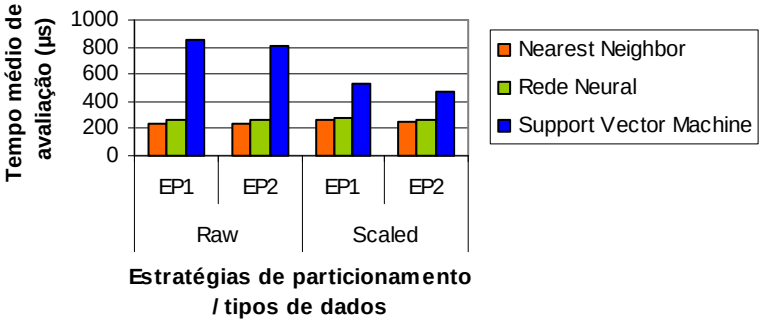

Figura 25: Tempos médios de avaliação – posturas.

7.2.2 Trajetórias

Em termos estruturais, a principal diferença entre as posturas e as trajetórias está em como são organizados os *frames* de dados que as

compõem. Todas as posturas seguem um padrão bem definido, no qual cinco valores representativos são utilizados. No caso das trajetórias, o número de pontos no espaço tridimensional (suas unidades básicas de composição) é variável; essa característica tem implicância direta na escolha dos classificadores a serem utilizados, e nas técnicas de pré-processamento que podem ser aplicadas para organizar melhor os dados disponíveis.

A primeira técnica escolhida para a classificação das trajetórias baseou-se em uma rede neural. É uma prática comum as redes neurais receberem como entrada o conjunto representativo completo de uma entidade a ser classificada. Para que essa prática pudesse ser adotada no presente trabalho, optou-se por igualar o tamanho de todas as trajetórias. O processo divide-se em quatro etapas, que ocorrem em cada iteração:

- particiona-se o conjunto de dados de acordo com o item 7.2 (para treinamento e avaliação);
- o conjunto de treinamento determina o tamanho das trajetórias que serão utilizadas. Para isso, é calculado o número de pontos médio das trajetórias que compõem o conjunto;
- todas as trajetórias do conjunto de treinamento têm seu tamanho igualado ao tamanho médio. Trajetórias que originalmente são menores incorporam novos pontos, distribuídos de forma equidistante pela sua extensão; trajetórias que originalmente são maiores têm os pontos excedentes excluídos, pontos esses selecionados também de forma equidistante. Após o ajuste de tamanhos, o conjunto é utilizado para o treinamento da rede neural;
- toda trajetória a ser avaliada passa pelo mesmo processo de ajuste de tamanho, sendo submetida em seguida ao classificador.

À primeira vista, o ajuste no tamanho de todas as trajetórias parece dispendioso em termos em processamento e tempo; uma análise com os dados coletados, porém, mostra que o processo é perfeitamente factível, inclusive para interfaces com restrições de tempo real. Na análise executada, chegou-se em um tempo médio de ajuste de 10,02ms por trajetória, com um desvio padrão de 5,08ms.

A rede neural configurada para a avaliação foi organizada em cinco camadas: uma camada de entrada com um número de nodos definido pelo tamanho médio das trajetórias, três camadas ocultas com o número de nodos igualmente definido pelo tamanho médio das trajetórias e uma camada de saída com sete nodos. A rede foi treinada

com uma taxa de aprendizado de 0,5, utilizando o algoritmo *backpropagation*. Foram escolhidas três estratégias de avaliação, com cada estratégia utilizando o melhor número de épocas (definido empiricamente a partir de treinamentos sucessivos, em um intervalo de 50 a 2.000 épocas) e um ajuste nos dados de treinamento e avaliação.

A primeira estratégia empregada foi executada com 200 épocas, treinando e avaliando a rede com as trajetórias ajustadas unicamente em termos de tamanho. Os resultados obtidos podem ser visualizados no Quadro 15, e indicam uma pequena diferença entre as estratégias de particionamento: 4,75% favorável à estratégia EP2.

Quadro 15: Método baseado em rede neural (com diversos pontos de origem).

		Classificado (EP1/EP2)							Precisão(%)
		T1	T2	T3	T4	T5	T6	T7	
Esperado (EP1/EP2)	T1	**609/539**	93/20	13/83	172/209	15/158	40/5	158/86	55,36/49,00
	T2	72/40	**404/316**	19/140	7/124	21/210	276/63	301/207	36,73/28,73
	T3	62/17	68/65	**594/850**	2/45	6/31	112/9	256/83	54,00/77,27
	T4	412/128	21/33	9/60	**473/688**	15/155	58/5	112/31	43,00/62,55
	T5	125/20	123/39	5/17	8/141	**595/833**	183/17	61/33	54,09/75,73
	T6	6/13	119/67	5/11	2/109	30/133	**871/722**	67/45	79,18/65,64
	T7	40/8	87/113	49/145	8/91	7/35	170/5	**739/703**	67,18/63,91
								Precisão média (%)	**55,65/60,40**

A segunda estratégia foi executada com 400 épocas, ajustando as trajetórias em relação a seus pontos de origem. As trajetórias são executadas livremente; porém, apesar de apresentarem traçados semelhantes, suas marcas de início e fim dificilmente coincidem em duas execuções (realizadas ou não pelo mesmo usuário). De forma a aumentar a similaridade dos padrões, as trajetórias de treinamento e avaliação foram ajustadas de forma a compartilharem o mesmo ponto de origem – as coordenadas (*0, 0, 0*). As coordenadas dos demais pontos de cada trajetória foram recalculadas, de forma a compensar os ajustes feitos em suas coordenadas iniciais. Os resultados obtidos (visualizados no Quadro 16) apresentam uma melhora considerável na acurácia do método.

Na terceira estratégia, além do ajuste das trajetórias em relação a suas coordenadas de origem, foi aplicada uma normalização dos valores das coordenadas para o intervalo [-1, 1]. A normalização visa reduzir os problemas numéricos citados no item 7.2.1. Os resultados obtidos (Quadro 17) diferem pouco daqueles observados no Quadro 16; uma explicação possível é de que o recálculo das coordenadas correspondente ao ajuste dos pontos de origem, por si só, reduz o intervalo de valores utilizados durante a classificação, levando a uma maior precisão.

Quadro 16: Método baseado em rede neural (com coordenadas de origem semelhantes).

		Classificado (EP1/EP2)							Precisão(%)
		T1	T2	T3	T4	T5	T6	T7	
Esperado (EP1/EP2)	T1	**992/996**	10/11	5/10	34/30	30/23	9/14	20/16	90,18/90,55
	T2	24/35	**978/964**	17/19	12/20	9/12	38/31	22/19	88,91/87,64
	T3	40/53	49/54	**945/933**	11/16	9/3	15/16	31/25	85,91/84,82
	T4	41/79	33/3	0/20	**970/950**	37/29	10/9	9/10	88,18/86,36
	T5	31/65	7/17	2/8	16/14	**994/929**	20/21	30/46	90,36/84,45
	T6	33/41	24/15	9/14	15/23	12/12	**982/969**	25/26	89,27/88,09
	T7	22/44	17/13	20/13	13/12	10/8	13/11	**1005/999**	91,36/90,82
								Precisão média (%)	**89,17/87,53**

Quadro 17: Método baseado em rede neural (700 épocas, com coordenadas de origem semelhantes e dados normalizados no intervalo [-1, 1]).

		Classificado (EP1/EP2)							Precisão(%)
		T1	T2	T3	T4	T5	T6	T7	
Esperado (EP1/EP2)	T1	**962/964**	13/8	6/11	27/39	67/50	8/10	17/18	87,45/87,64
	T2	26/27	**927/983**	11/9	9/3	56/47	51/28	20/3	84,27/89,36
	T3	66/96	42/37	**928/901**	4/7	29/32	6/13	25/14	84,36/81,91
	T4	84/72	9/7	8/2	**937/955**	53/48	0/5	9/11	85,18/86,82
	T5	27/19	7/8	0/3	8/19	**1043/1024**	9/9	6/18	94,82/93,09
	T6	36/38	35/34	5/7	15/14	38/34	**949/964**	22/9	86,27/87,64
	T7	22/25	32/27	6/16	9/9	31/36	29/31	**971/956**	88,27/86,91
								Precisão média (%)	**87,23/87,62**

A avaliação dos métodos teve continuidade com a heurística baseada em AFDs. Diferentemente da rede neural, os autômatos construídos não recebem como entrada uma trajetória completa, mas a sequência de pontos (um de cada vez) que compõem a trajetória a ser classificada. Outra diferença está no número de classificadores: a rede neural é única, convergindo para um dentre sete resultados possíveis; no caso dos autômatos, é possível construir um classificador para cada padrão de treinamento (em outras palavras, um autômato para cada trajetória de treinamento), ou agrupar um conjunto de trajetórias em uma *trajetória média*, que por sua vez é utilizada na construção do classificador. No momento da avaliação, o autômato que apresentar o estado atual mais avançado (ou que tiver atingido seu estado final) define o resultado do reconhecimento.

Os primeiros resultados da heurística baseada em AFDs foram obtidos utilizando-se como dados de entrada as trajetórias com seus tamanhos originais, porém com seus pontos de origem ajustados para as coordenadas (*0*, *0*, *0*). Mais uma vez, a estratégia de particionamento EP1 mostrou-se mais adequada como regra de separação de conjuntos de treinamento/avaliação, com um ganho equivalente a 7,6% em relação à estratégia EP2. O Quadro 18 exibe os resultados completos, considerando uma tolerância de 15º para os ângulos φ e θ.

Quadro 18: Método baseado em AFDs (trajetórias originais, com coordenadas de origem semelhantes e tolerância de 15º).

		Classificado (EP1/EP2)							Precisão(%)
		T1	T2	T3	T4	T5	T6	T7	
Esperado (EP1/EP2)	T1	**599/360**	29/27	367/545	72/98	2/1	31/17	0/52	54,45/32,73
	T2	49/13	**905/639**	42/35	18/13	45/149	27/79	14/172	82,27/58,09
	T3	7/6	92/109	**775/734**	11/7	119/89	27/41	69/114	70,45/66,73
	T4	97/81	12/15	92/85	**872/867**	2/0	25/11	0/41	79,27/78,82
	T5	59/20	50/56	111/129	14/5	**811/739**	9/45	46/106	73,73/67,18
	T6	10/6	129/162	140/187	14/15	26/9	**632/556**	149/165	57,45/50,55
	T7	22/25	114/72	34/14	2/1	206/123	244/273	**478/592**	43,45/53,82
								Precisão média (%)	**65,87/58,27**

Para a continuidade da avaliação da heurística baseada em AFDs, as trajetórias foram igualadas em tamanho, utilizando-se o mesmo processo de quatro etapas descrito para a rede neural (mantendo-se as características pertinentes a cada classificador). De forma similar, os resultados obtidos foram melhores – como pode ser visualizado no Quadro 19.

Quadro 19: Método baseado em AFDs (trajetórias com mesmo tamanho, coordenadas de origem semelhantes e tolerância de 15º).

		Classificado (EP1/EP2)							Precisão(%)
		T1	T2	T3	T4	T5	T6	T7	
Esperado (EP1/EP2)	T1	**800/720**	8/13	19/28	161/168	75/63	25/101	12/7	72,73/65,45
	T2	58/50	**961/794**	5/33	10/17	2/3	21/97	43/106	87,36/72,18
	T3	30/27	26/15	**681/709**	15/30	120/118	34/74	194/127	61,91/64,45
	T4	94/106	2/5	25/22	**934/911**	1/1	34/53	10/2	84,91/82,82
	T5	114/62	63/62	13/20	4/5	**768/796**	72/76	66/79	69,82/72,36
	T6	46/29	111/89	164/73	8/16	51/62	**610/692**	110/139	55,45/62,91
	T7	57/27	47/33	8/7	0/1	87/91	82/125	**819/816**	74,45/74,18
								Precisão média (%)	**72,38/70,62**

Como última estratégia de avaliação para a heurística, as trajetórias de treinamento foram agrupadas por classe. Esse agrupamento gera, no lugar de 385 padrões de treinamento, apenas sete padrões representativos. O maior ganho com a adoção da estratégia está no tempo total de avaliação de uma trajetória, que é reduzido drasticamente devido à redução massiva do número de autômatos cujas funções de transição devem ser executadas. Em termos de precisão, os resultados praticamente equivalem àqueles obtidos com as trajetórias originais (a primeira estratégia testada para o método). O Quadro 20 exibe a matriz de confusão obtida com uma tolerância de 20º para os ângulos φ e θ.

Quadro 20: Método baseado em AFDs (trajetórias agrupadas, coordenadas de origem semelhantes e tolerância de 20º).

		Classificado (EP1/EP2)							Precisão(%)
		T1	T2	T3	T4	T5	T6	T7	
Esperado (EP1/EP2)	T1	**0/20**	83/91	291/264	342/302	101/128	215/203	68/92	0,00/1,82
	T2	0/26	**966/917**	27/45	35/30	3/1	15/42	54/39	87,82/83,36
	T3	0/20	75/72	**762/739**	21/24	97/112	79/81	66/52	69,27/67,18
	T4	0/17	13/27	31/33	**1019/1005**	4/4	7/4	26/10	92,64/91,36
	T5	0/20	151/160	6/26	20/23	**773/730**	118/118	32/23	70,27/66,36
	T6	0/21	223/215	71/87	9/11	77/83	**647/613**	73/70	58,82/55,73
	T7	0/20	137/128	18/24	0/1	59/62	369/336	**517/529**	47,00/48,09
								Precisão média (%)	**60,83/59,13**

A avaliação foi concluída com o método baseado em modelos ocultos de Markov. Essa avaliação não utilizou apenas um HMM, mas uma rede de HMMs representando as sete classes de trajetórias que compõem o vocabulário de gestos. De forma semelhante ao método baseado em autômatos, os HMMs recebem como entrada as coordenadas dos pontos componentes das trajetórias, convergindo para o resultado mais provável. Duas estratégias foram escolhidas para a validação do método: na primeira, as trajetórias originais (sem ajuste de tamanho), com coordenadas de origem semelhantes, foram avaliadas; na segunda, foram utilizadas trajetórias com mesmo tamanho e coordenadas de origem semelhantes. Os modelos foram construídos com, respectivamente, oito e sete estados, adotando a topologia *left-right* e treinados com o algoritmo *Baum-Welch*. Os resultados obtidos foram compilados nos Quadros 21 e 22, e os melhores resultados para as trajetórias podem ser visualizados na Fig. 26.

Quadro 21: Método baseado em HMMs (trajetórias originais, com coordenadas de origem semelhantes e 8 estados).

		Classificado (EP1/EP2)							Precisão(%)
		T1	T2	T3	T4	T5	T6	T7	
Esperado (EP1/EP2)	T1	**524/327**	93/39	254/491	72/93	43/19	84/73	30/58	47,64/29,73
	T2	66/41	**817/602**	42/74	35/43	58/112	47/89	35/139	74,27/54,73
	T3	55/26	82/109	**755/692**	11/15	101/89	27/50	69/119	68,64/62,91
	T4	79/79	12/56	87/85	**770/752**	52/55	87/22	13/51	70,00/68,36
	T5	66/45	56/58	98/124	28/45	**756/678**	47/57	49/93	68,73/61,64
	T6	17/63	130/160	114/187	22/27	26/18	**689/480**	102/165	62,64/43,64
	T7	12/32	102/80	51/32	49/18	186/129	213/269	**487/540**	44,27/49,09
								Precisão média (%)	**62,31/52,87**

Quadro 22: Método baseado em HMMs (trajetórias com mesmo tamanho, com coordenadas de origem semelhantes e 7 estados).

		Classificado (EP1/EP2)							Precisão(%)
		T1	T2	T3	T4	T5	T6	T7	
Esperado (EP1/EP2)	T1	**713/676**	74/22	25/11	132/163	75/48	56/145	25/35	64,82/61,45
	T2	61/68	**854/715**	46/44	72/21	12/45	21/112	34/95	77,64/65,00
	T3	32/45	33/29	**650/622**	27/29	99/108	56/142	203/125	59,09/56,55
	T4	72/92	57/34	32/59	**869/820**	11/23	14/43	45/29	79,00/74,55
	T5	128/84	82/53	13/59	4/41	**702/711**	82/70	89/82	63,82/64,64
	T6	58/78	99/92	182/73	27/44	51/62	**590/599**	93/152	53,64/54,45
	T7	37/25	86/46	21/24	32/65	77/81	72/103	**775/756**	70,45/68,73
								Precisão média (%)	**66,92/63,62**

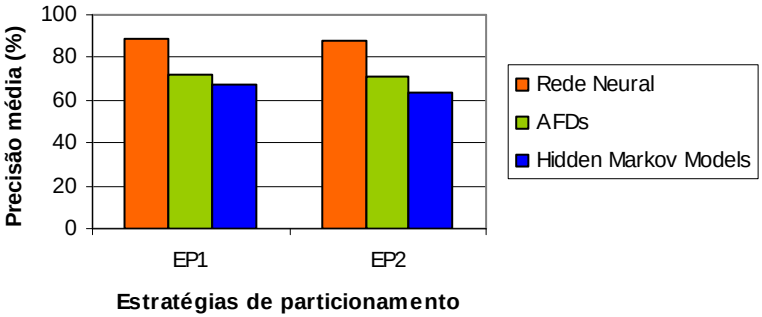

Figura 26: Melhores resultados obtidos na avaliação das trajetórias.

O desempenho dos métodos e estratégias selecionados pode ser visualizado no Quadro 23. A tabela relaciona os métodos com seus respectivos resultados, consolidando as estratégias de particionamento. Convém lembrar que, no caso do método baseado em AFDs, o tempo de treinamento corresponde ao tempo de construção dos classificadores. A diferença no desempenho dos métodos pode ser visualizada na Fig. 27.

Pela análise dos dados apresentados neste capítulo, é possível identificar alguns elementos responsáveis pelas maiores ou menores taxas de reconhecimento obtidas pelos classificadores. O capítulo a seguir apresenta esses elementos, explicitando a sua relevância e sua relação com essas taxas.

Quadro 23: Desempenho das etapas de treinamento e avaliação – trajetórias.

	Desempenho			
	Trajetórias originais		Trajetórias com mesmo tamanho	
	TMT/DP	TMA/DP	TMT/DP	TMA/DP
Rede Neural	-	-	25,85min/11,42min	2,55ms/0,17ms
AFDs	4,22s/0,14s	1,06s/0,42s	4,23s/0,06s	1,12s/0,03s
HMMs	17,10s/1,00s	7,07ms/2,18ms	16,98s/1,02s	6,96ms/2,15ms

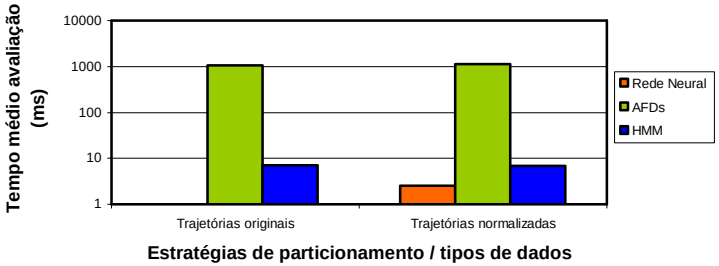

Figura 27: Tempos médios de avaliação – trajetórias.

8. Discussão

Os resultados apresentados no capítulo anterior resumem, em termos de precisão e desempenho, o comportamento dos métodos de reconhecimento de gestos selecionados para este trabalho. O presente capítulo discute esses resultados, subsidiando as conclusões finais relatadas no capítulo nove.

Inicialmente, uma análise dos resultados obtidos permite atestar a robustez do ambiente experimental utilizado para a aquisição dos dados de treinamento e validação. A dinâmica obtida com os vídeos de exemplo dos gestos a serem mimetizados mostrou-se eficiente em termos de entendimento, e o *feedback* visual imediato dado aos usuários por meio da *webcam* mostrou-se efetivo no auxílio à correção das posturas e trajetórias executadas. Outro fator que contribuiu para o sucesso do ambiente experimental foi a presença de um especialista durante a primeira iteração executada pelos usuários, guiando-os pelas etapas do processo de coleta de dados.

As estratégias de particionamento adotadas no trabalho (EP1 e EP2) mostraram-se relevantes em relação aos resultados obtidos. A heterogeneidade dos dados resultantes da aplicação da estratégia EP1 garantiu uma melhor generalização por parte dos métodos selecionados, reduzindo os erros sistemáticos gerados pela adoção de dados provenientes de um conjunto limitado de usuários. Apesar de o mesmo usuário dificilmente executar o mesmo gesto de forma idêntica, há uma tendência de que este gere padrões bastante similares, o que limita a diversidade do conjunto de treinamento e influencia nas classificações.

Para as posturas, a escolha dos dados do tipo *raw* e *scaled* também influencia na acurácia dos reconhecedores. Como pode ser visualizado nas matrizes de confusão, dados *scaled* apresentam melhores taxas de reconhecimento, o que justifica a adoção de um procedimento de calibragem a ser executado por cada um dos usuários envolvidos. Essa calibragem acaba por gerar uma *identidade anatômica* para os usuários, que pode ser armazenada e recuperada sempre que necessário. Outro fator importante a ser considerado diz respeito às implicações da definição do vocabulário de gestos em relação ao *hardware* adotado, e vice-versa. As posturas rotuladas como P02, P03 e P04, bem como as posturas rotuladas como P11 e P14, apresentam uma identidade visual distinta, facilmente identificável por métodos de visão computacional; porém, seus padrões instrumentados são bastante semelhantes e difíceis de discernir. A alternativa, nesses casos, implica em uma adaptação a ser feita no vocabulário (com o agrupamento de

gestos semelhantes) ou no *hardware* (adotando-se equipamentos capazes de gerar dados suficientes para uma diferenciação).

Para as trajetórias, percebe-se que a normalização em termos de tamanho e origem influencia positivamente na acurácia de classificação. Igualar trajetórias em tamanho permite atender às restrições de todos os classificadores escolhidos, principalmente aqueles que exigem uma entrada de tamanho fixo; normalizar trajetórias a partir de suas coordenadas de origem permite organizar a orientação espacial das mesmas, tornando-as menos esparsas e, consequentemente, mais similares. O tempo de normalização computado durante os testes mostrou-se aceitável (mesmo para interfaces com restrições de tempo real), o que torna a adaptação dos dados um processo factível.

Finalmente, o desempenho computado para a etapa de avaliação permite a utilização de qualquer um dos métodos selecionados em ambientes interativos (desde que sejam adotadas as estratégias de construção e configuração descritas no presente trabalho). Os tempos de treinamento, por sua vez, correspondem à primeira execução dos classificadores; os resultados dessa etapa serão sempre os mesmos – desde que os padrões de treinamento sejam mantidos, em termos de volume; assim, os classificadores que dependem dessa etapa podem ser construídos com base em treinamentos feitos anteriormente, o que reduz o tempo de *startup* dos métodos mantendo sua acurácia.

Com base nos resultados apresentados no capítulo sete e na identificação dos principais elementos responsáveis pela obtenção desses resultados neste capítulo, o capítulo a seguir relata as conclusões do presente trabalho quanto à escolha dos métodos de reconhecimento de gestos adequados, aplicáveis a interfaces instrumentadas.

9. Conclusões e trabalhos futuros

Neste trabalho foi realizada uma comparação de técnicas de reconhecimento de posturas e trajetórias, objetivando identificar os melhores métodos a serem utilizados na construção de interfaces gestuais baseadas em dados instrumentados. Os dados para a comparação foram coletados a partir de um conjunto de usuários voluntários, através de uma aplicação desenvolvida especificamente com essa finalidade. Foram testadas diferentes estratégias de construção, treinamento e configuração das técnicas escolhidas, de forma a identificar os elementos que efetivamente influenciam em seus resultados. Ficou claro que a escolha de uma técnica específica depende não apenas de uma característica, mas de um conjunto de características complementares que, somadas, determinam o comportamento de cada classificador.

Dentre os elementos relevantes identificados, destaca-se a estratégia de particionamento (provando que conjuntos de treinamento heterogêneos melhoram a generalização atingida pelos classificadores). Especificamente para as posturas, a importância da calibragem e da relação entre o vocabulário e o *hardware* disponível merece uma atenção especial; a primeira por adaptar os dados à anatomia particular de cada usuário, e a segunda por identificar que padrões instrumentados podem estabelecer limites quanto à acurácia dos reconhecimentos, se capturados com dispositivos mais limitados. Para as trajetórias, percebeu-se que o ajuste dos tamanhos de acordo com um padrão e a sua orientação no espaço são fatores importantes e que devem ser observados; ambos auxiliam na restrição espacial dos gestos, o que contribui para agrupar trajetórias pertencentes a uma mesma classe dada a sua maior similaridade.

Em termos de precisão, os melhores resultados para posturas foram obtidos pelo método baseado em SVM, com 79,13% e 90,96% de reconhecimentos positivos, respectivamente, para o vocabulário original e o vocabulário com gestos agrupados. As trajetórias, por sua vez, foram melhor classificadas pela rede neural, com uma taxa de reconhecimento de 89,17%. Quanto ao desempenho, as avaliações mais rápidas para posturas foram executadas pelo método baseado em *nearest neighbor* (média de 229,82μs), e para trajetórias pelo método baseado em rede neural (média de 2,55ms). Dados os baixos tempos de avaliação atingidos pela totalidade dos métodos testados, o principal critério de escolha passou a ser a precisão obtida, que varia consideravelmente entre as técnicas. Assim, a principal conclusão da presente dissertação é de que a combinação SVM/rede neural, respectivamente aplicada ao

reconhecimento de posturas e trajetórias, é a melhor escolha quando o cenário de aplicação envolve dados instrumentados, passíveis de normalização e com estratégias de partição que garantam heterogeneidade para os conjuntos de treinamento.

Como proposta de trabalhos futuros, sugere-se:

- a validação dos melhores métodos identificados nesta dissertação através da implementação de uma interface gestual, objetivando verificar se as taxas de reconhecimento obtidas são suficientes para um ambiente de execução real. Como aplicação-alvo, propõe-se a utilização do *framework* 3D desenvolvido pelo grupo Cyclops, da Universidade Federal de Santa Catarina (UFSC), organizado com o intuito de prover visualização e manipulação de dados médicos em ambientes tridimensionais (TANI et al., 2006 e SILVA et al., 2009);
- a utilização dos dados adquiridos para este trabalho (disponíveis publicamente em http://www.lapix.ufsc.br/index.php/gil-gesture-interaction-layer) na validação de outras técnicas, tanto instrumentadas quanto baseadas em visão computacional;
- o aperfeiçoamento da heurística baseada em autômatos, que em termos de precisão ficou à frente dos modelos de Markov. Dada a simplicidade do modelo proposto, diferentes funções de transição podem ser testadas, melhorando ainda mais seus resultados; quanto ao desempenho, a heurística pode ser adaptada para executar as funções de transição em paralelo, explorando recursos de ambientes *multicore* ou mesmo GPUs (*Graphics Processing Units*);
- o aprimoramento da aplicação desenvolvida para a coleta de dados, com seu posterior reuso em estudos voltados à interpretação de gestos livres.

Referências bibliográficas

ACM SIGCHI. **Curricula for Human-Computer Interaction.** Disponível em: http://old.sigchi.org/cdg/. Acessado em: 19 de Janeiro de 2009.

ALLEN, B. D., BISHOP, G., WELCH, G. **Tracking: Beyond 15 Minutes of Thought.** Annual Conference on Computer Graphics & Interactive Techniques, course 11, 2001.

ARGYROS, A. A., LOURAKIS, M. I. A. **Vision-Based Interpretation of Hand Gestures for Remote Control of a Computer Mouse.** Computer Vision in Human-Computer Interaction, p. 40-51, 2006.

BAILADOR, G., ROGGEN, D., TRÖSTER, G. et al. **Real time gesture recognition using Continuous Time Recurrent Neural Networks.** Proceedings of the ICST 2nd International Conference on Body Area Networks, p. 1-8, 2007.

BOWMAN, D. A., KRUIJFF, E., LAVIOLA, J. J. et al. **3D User Interfaces: Theory and Practice.** Boston: Addison-Wesley, 2004.

BREUER, P., ECKES, C., MÜLLER, S. **Hand Gesture Recognition with a Novel IR Time-of-Flight Range Camera–A Pilot Study.** Lecture Notes in Computer Science, v. 4418, p. 247-260, 2007.

CARIDAKIS, G., KARPOUZIS, K., PATERITSAS, C. et al. **Hand Trajectory Based Gesture Recognition Using Self-Organizing Feature Maps and Markov Models.** Proceedings of the IEEE International Conference on Multimedia and Expo, p. 1105-1108, 2008.

CHEN, Y-T., TSENG, K-T. **Developing a Multiple-Angle Hand Gesture Recognition System for Human Machine Interactions.** The 33rd Annual Conference of the IEEE Industrial Electronics Society, p. 489-492, 2007.

DELLER, M., EBERT, A., BENDER, M. et al. **Flexible Gesture Recognition for Immersive Virtual Environments.** Tenth International Conference on Information Visualization, p. 563-568, 2006.

DIPIETRO, L., SABATINI, A. M., DARIO, P. **A Survey of Glove-Based Systems and Their Applications.** IEEE Transactions on Systems, Man, and Cybernetics, Part C, v. 38, p. 461-482, 2008.

ELMEZAIN, M., AL-HAMADI, A., MICHAELIS, B. **Real-Time Capable System for Hand Gesture Recognition Using Hidden**

Markov Models in Stereo Color Image Sequences. Journal of WSCG, v. 16, n. 1-3, p. 65-72, 2008.

FANG, Y., CHENG, J., WANG, K. et al. **Hand Gesture Recognition Using Fast Multi-scale Analysis.** Proceedings of the Fourth International Conference on Image and Graphics, p. 694-698, 2007.

FOXLIN, E. **Motion Tracking Requirements and Technologies.** In STANEY, K. Handbook of Virtual Environment Technology, 2002.

FROLOV, V., DEML, B., HANNIG, G. **Gesture Recognition with Hidden Markov Models to Enable Multi-modal Haptic Feedback.** Proceedings of the 6th International Conference on Haptics: Perception, Devices and Scenarios, p. 786-795, 2008.

GABBARD, J. L., HIX, D., SWAN, J. E. **User-Centered Design and Evaluation of Virtual Environments.** IEEE Computer Graphics and Applications, v. 19, n. 6, p. 51-59, 1999.

GABBARD, J. L., HIX, D., SWAN, J. E. et al. **Usability Engineering for Complex Interactive Systems Development.** Proceedings of Human Systems Integration Symposium 2003, Engineering for Usability, p. 1-13, 2003.

GORDON, G., CHEN, X., BUCK, R. **Person and Gesture Tracking with Smart Stereo Cameras.** Proceedings of SPIE: Three-Dimensional Image Capture and Applications, v. 6805, p. 68050T.1-68050T.11, 2008.

HARTSON, H. R., ANDRE, T. S., WILLIGES, R. W. **Criteria for Evaluating Usability Evaluation Methods.** International Journal of Human-Computer Interaction, v. 15, n. 1, p. 145-181, 2003.

HIX, D., HARTSON, H. R. **Formative Evaluation: Ensuring Usability in User Interfaces.** Technical Report. TR-92-60, Virginia Polytechnic Institute & State University, 1992.

HOPCROFT, J. E., MOTWANI, J., ULLMAN, J. D. Introduction to Automata Theory, Languages, and Computation. Boston: Addison-Wesley, 2001.

ISHIKAWA, M., MATSUMURA, H. **Recognition of a Hand-Gesture Based on Self-Organization Using a DataGlove.** Proceedings of the 6th International Conference on Neural Information Processing, v. 2, p. 739-745, 1999.

IWAI, Y., SHIMIZU, H., YACHIDA, M. **Real-time Context-based Gesture Recognition Using HMM and Automaton.** Proceedings on International Workshop on Recognition,

Analysis, and Tracking of Faces and Gestures in Real-Time Systems, p. 127-134, 1999.

JUST, A., RODRIGUEZ, Y., MARCEL, S. **Hand Posture Classification and Recognition using the Modified Census Transform.** Proceedings of the 7th International Conference on Automatic Face and Gesture Recognition, p. 351-356, 2006.

KARAM, M., SCHRAEFEL, M. C. **A Taxonomy of Gestures in Human Computer Interaction.** Technical Report. University of Southampton, 2005.

KELLEY, J. F. **An Iterative Design Methodology for User-Friendly Natural Language Office Information Applications.** ACM Transactions on Information Systems (TOIS), v. 2, n. 1, p. 26-41, 1984.

KJELLSTRÖM, H., ROMERO, J., KRAGIĆ, D. **Visual Recognition of Grasps for Human-to-Robot Mapping.** IEEE/RSJ International Conference on Intelligent Robots and Systems, p. 3192-3199, 2008.

LAVIOLA, J. J. **A Survey of Hand Posture and Gesture Recognition Techniques and Technology.** Technical Report. UMI Order Number: CS-99-11. Brown University, 1999.

LEE, C., GHYME, S., PARK, C. et al. **The Control of Avatar Motion Using Hand Gesture.** Proceedings of the ACM Symposium on Virtual Reality Software and Technology, p. 59-65, 1998.

LIU, Y., GAN, Z., SUN, Y. **Static Hand Gesture Recognition and Its Application based on Support Vector Machines.** Proceedings of the 9th ACIS International Conference on Software Engineering, Artificial Intelligence, Networking, and Parallel/Distributed Computing, p. 517-521, 2008.

MALIMA, A., ÖZGÜR, E., ÇETIN, M. **A Fast Algorithm for Vision-Based Hand Gesture Recognition for Robot Control.** Proceedings of the 14th IEEE Signal Processing and Communications Applications, p. 1-4, 2006.

MANRESA, C., VARONA, J., MAS, R. et al. **Hand Tracking and Gesture Recognition for Human-Computer Interaction.** Electronic Letters on Computer Vision and Image Analysis, v. 5, n. 3, p. 96-104, 2005.

MARCEL, S. **Evaluation Protocols and Comparative Results for the Triesch Hand Posture Database.** Research Report. IDIAP-RR 02-50, Dalle Molle Institute for Perceptual Artificial Intelligence, 2002.

MARCEL, S. **Hand Posture and Gesture Datasets.**

Disponível em: http://www.idiap.ch/resources/gestures/. Acessado em: 20 de Fevereiro de 2009.

MAZZA, R. **Evaluating Information Visualization Applications with Focus Groups: the CourseVis Experience.** Proceedings of the 2006 AVI Workshop on Beyond Time and Errors: Novel Evaluation Methods for Information Visualization, p. 1-6, 2006.

MENG, H., PEARS, N., BAILEY, C. **A Human Action Recognition System for Embedded Computer Vision Application.** IEEE Conference on Computer Vision and Pattern Recognition, p. 1-6, 2007.

MERRIAM-WEBSTER. **Gesture.** Disponível em: http://www.merriam-webster.com/dictionary/gesture. Acessado em: 10 de Fevereiro de 2009.

MITRA, S., ACHARYA, T. **Gesture Recognition: A Survey.** IEEE Transactions on Systems, Man, and Cybernetics, Part C, v. 37, p. 311-324, 2007.

MYERS, B. A. **A Brief History of Human-Computer Interaction Technology.** ACM interactions, v. 5, n. 2, p. 44-54, 1998.

NEHANIV, C. L. **Classifying Types of Gesture and Inferring Intent.** Proceedings of the Symposium on Robot Companions: Hard Problems and Open Challenges in Robot-Human Interaction, 2005a.

NEHANIV, C. L., DAUTENHAHN, K., KUBACKI, J. et al. **A methodological approach relating the classification of gesture to identification of human intent in the context of human-robot interaction.** IEEE International Workshop on Robot and Human Interactive Communication, p. 371-377, 2005b.

PAVLOVIĆ, V. I., SHARMA, R., HUANG, T. S. **Visual Interpretation of Hand Gestures for Human-Computer Interaction: A Review.** IEEE Transactions on Pattern Analysis and Machine Intelligence, v. 19, n. 7, p. 677-695, 1997.

POPPE, R. **Vision-based human motion analysis: An overview.** Computer Vision and Image Understanding, v. 108, n. 1-2, p. 4-18, 2007.

REN, Y., ZHANG, F. **Hand Gesture Recognition Based on MEB-SVM.** Proceedings of the 2009 International Conference on Embedded Software and Systems, p. 344-349, 2009.

RIZZO, A. A., KIM, G. J., YEH, S-C. et al. **Development of a Benchmarking Scenario for Testing 3D User Interface Devices and Interaction Methods.** Proceedings of the 11th International Conference on Human Computer Interaction, 2005.

ROLLAND, J. P., BAILLOT, Y., GOON, A. A. **A survey of tracking technology for virtual environments.** In BARFIELD, W., CAUDELL, T. Fundamentals of Wearable Computers and Augmented Reality, p. 67-112, 2001.

SARLE, W. S. **Neural Network FAQ.** Disponível em: ftp://ftp.sas.com/pub/neural/FAQ.html. Acessado em: 18 de Outubro de 2009.

SCHLÖMER, T., POPPINGA, B., HENZE, N. et al. **Gesture Recognition with a Wii Controller.** Proceedings of the 2nd International Conference on Tangible and Embedded Interaction, p. 11-14, 2008.

SILVA, A. F. B., NOBREGA, T. H. C., CARVALHO, D. D. B. et al. **Framework for Interactive Medical Imaging Applications.** Colloquium of Computation: Brazil / INRIA, Cooperations, Advances and Challenges, p. 126-129, 2009.

STEFAN, A., ATHITSOS, V., ALON, J. et al. **Translation and Scale-Invariant Gesture Recognition in Complex Scenes.** Proceedings of the 1st International Conference on Pervasive Technologies Related to Assistive Environments, p. 1-8, 2008.

STERGIOPOULOU, E., PAPAMARKOS, N. **Hand gesture recognition using a neural network shape fitting technique.** Engineering Applications of Artificial Intelligence, v. 22, n. 8, p. 1141-1158, 2009.

TANI, B. S., NOBREGA, T., SANTOS, T. R. et al. **Generic Visualization and Manipulation Framework for Three-Dimensional Medical Environments.** Proceedings of the 19th IEEE International Symposium on Computer-Based Medical Systems, p. 27-31, 2006.

TANI, B. S., MAIA, R. S., WANGENHEIM, A. v. **A Gesture Interface for Radiological Workstations.** Proceedings of the 20th IEEE International Symposium on Computer-Based Medical Systems, p. 27-32, 2007.

TARRATACA, L., SANTOS, A. C., CARDOSO, J. M. P. **The Current Feasibility of Gesture Recognition for a Smartphone using J2ME.** Proceedings of the 2009 ACM Symposium on Applied Computing, p. 1642-1649, 2009.

TRIESCH, J., MALSBURG, C. v. d. **Robust Classification of Hand Postures Against Complex Backgrounds.** Proceedings of the 2nd International Conference on Automatic Face and Gesture Recognition, p. 170-175, 1996.

WATSON, R. **A Survey of Gesture Recognition Techniques.** Technical Report. TCD-CS-93-11, Trinity College, Dublin, 1993.

WEISSMANN, J., SALOMON, R. **Gesture Recognition for Virtual Reality Applications Using Data Gloves and Neural Networks.** International Joint Conference on Neural Networks, v. 3, p. 2043-2046, 1999.

WELCH, G., FOXLIN, E. **Motion Tracking: No Silver Bullet, but a Respectable Arsenal.** IEEE Computer Graphics and Applications, v. 22, n. 6, p. 24-38, 2002.

WEXELBLAT, A. D. **A Feature-Based Approach to Continuous-Gesture Analysis.** Tese de mestrado. Massachusetts: Massachusetts Institute of Technology, 1994.

WOBBROCK, J. O., MORRIS, M. R., WILSON, A. D. **User-defined gestures for surface computing.** Proceedings of the 27th international conference on Human factors in computing systems, p. 1083-1092, 2009.

XU, D., YAO, W., ZHANG, Y. **Hand Gesture Interaction for Virtual Training of SPG.** International Conference on Artificial Reality and Telexistence – Workshops, p. 672-676, 2006.

YEUNG, C-H., LAM, M-W., CHAN, H-C. et al. **Vision-Based Hand Gesture Interactions for Large LCD-TV Display Tabletop Systems.** Proceedings of the 9th Pacific Rim Conference on Multimedia: Advances in Multimedia Information Processing, p. 89-98, 2008.

ZIAIE, P., MÜLLER, T., FOSTER, M. E. et al. **Using a Naïve Bayes Classifier based on K-Nearest Neighbors with Distance Weighting for Static Hand-Gesture Recognition in a Human-Robot Dialog System.** Proceedings of the 13th International CSI Computer Conference, 2008.

Apêndice A – Publicações

A.1 Comparative Evaluation of Static Gesture Recognition Techniques based on Nearest Neighbor, Neural Networks and Support Vector Machines

Tipo: Full-Paper

Qualificação: B2

Autores: Alexandre Savaris e Aldo von Wangenheim

Periódico: Journal of the Brazilian Computer Society

Status: Aceito para publicação