Uma Ferramenta para Construção de Conjuntos de Dados de Referência para Sistemas de Análise de Gestos Baseados em Imagens

Priscilla K. Wagner, Guilherme O. Borges, Renata C. B. Madeo, Sarajane M. Peres[1]

[1]Escola de Artes, Ciências e Humanidades - Universidade de São Paulo (USP)
Av. Arlindo Béttio – 03828-000 – São Paulo – SP – Brasil

{priscilla.wagner, guilherme.borges, renata.si, sarajane}@usp.br

Abstract. *This paper presents a tool for building benchmark datasets to evaluate image-based gesture analysis systems, considering object segmentation and pattern recognition tasks. The main motivation in this project is to provide an alternative to allow the efective comparisons among several iniciatives in the development of this type of systems. The tool has as a differential funcionalities for preparing dataset with videos, frames and feature vectors organized according acquision context, background style, performers, gesture sequences and gesture type. Such features allow to create test cases to benchmarking several systems properties.*

Resumo. *Este artigo apresenta uma ferramenta para construção de conjuntos de dados de referência para aferição de sistemas de análise de gestos baseados em imagens, visando tarefas de segmentação de objetos de interesse e reconhecimento de padrões. A principal motivação deste projeto é contribuir para que as várias iniciativas de desenvolvimento de tais sistemas possam ser eficazmente comparadas. A ferramenta apresenta como diferencial funcionalidades dedicadas a preparar conjuntos de vídeos, frames e vetores de características, organizados em relação ao contexto de aquisição, tipo de fundo, usuário executor, sequências e tipos de gestos. Estas características permitem a criação de casos de testes sob os quais várias propriedades do sistema podem ser aferidas.*

1. Introdução

Sistemas para análise de gestos baseados em imagens tem se tornado cada vez mais comuns nos dias atuais. As áreas de aplicação para estes sistemas vão desde reconhecimento de face para sistemas de identificação de pessoas em fotos postadas em redes sociais ou reconhecimento de sorrisos em sistemas embarcados em máquinas fotográficas, até sistemas de segurança, reconhecimento de língua de sinais e suporte à análise de discurso em atividades de comunicação humana. A informação referente a gestos estudada neste tipo de sistema assume características estáticas, como por exemplo no reconhecimento de face [Pantic et al. 2007], [Ekman e Hager 2002] ou reconhecimento de postura corporal [Just et al. 2006] [Madeo et al. 2010]; ou dinâmicas como no caso de análise de movimento corporal [Dias et al. 2009] ou direção do olhar [Duchowski 2002].

O campo de pesquisa nesta área se apresenta como multidisciplinar, envolvendo pesquisadores das áreas de desenvolvimento de sistemas, fisiologia, psicologia e linguística. Mitra e Acharya [Mitra e Acharya 2007] revisaram uma série de métodos

computacionais, propostos por diversos pesquisadores, para análise de gestos a partir de imagens (também conhecido como análise de gestos baseado em visão), e Kita [Kita et al. 1998], Kendon [Kendon 1980] e McNeil [McNeill 1992] apresentam estudos de psicologia e linguística que buscam fundamentação teórica para formalizar a análise de gestos. Tal formalização é imprescindível para que a análise de gestos computacionais alcance seu pleno desenvolvimento.

Porém, ainda mediante os esforços para crescimento e maturação desta área de pesquisa e desenvolvimento, permanece a dificuldade de realização de análises comparativas entre as abordagens propostas. Por conta do caráter cultural e regional inerente ao próprio problema de análise de gestos, e a depender da aplicação, o objeto de estudo – o gesto – tem significados distintos quando contextualizado em diferentes países ou mesmo em contextos específicos de aplicações particulares. Existe uma carência de conjuntos de dados de gestos – bem descritos e organizados e que sejam de domínio público – para serem usados como conjuntos de referência (ou *benchmark*) para aferição de desempenho e testes comparativos entre as iniciativas.

Criar conjuntos de dados de referência nesta área pode ser uma tarefa desgastante, uma vez que o ideal é que os dados que componham o conjunto sejam em grande quantidade e diversidade, principalmente quando uma técnica computacional para segmentação de imagem ou para reconhecimento de padrões é escolhida para a realização da análise do gesto, como é o caso do escopo coberto pelo presente trabalho[1]. A ferramenta aqui descrita, desenvolvida usando o ambiente de programação Matlab, tem o objetivo de contribuir para a minimização deste problema, fornecendo um ambiente para aquisição e organização de dados para compor tais conjuntos de referências.

Apesar da pouca disponibilidade de conjuntos de dados de referência na área de gestos, é possível citar alguns conjuntos relacionados a línguas de sinais disponíveis publicamente no repositório UCI Machine Learning[2]. O conjunto Australian Sign Language (Língua de Sinais Australiana), gerado por [Kadous 1995], inclui 6650 instâncias de 95 sinais distintos da Língua de Sinais Australiana (70 instâncias por classe, em média), porém os sinais foram extraídos utilizando luvas de sensores. Já o conjunto Libras Movement (Movimentos da Língua Brasileira de Sinais – Libras), apresentado em [Dias et al. 2009] consiste em 360 instâncias de 15 movimentos distintos (24 instâncias por classe, em média). O trabalho utiliza visão computacional, de forma que os dados são captados na forma de vídeos, sendo necessário extrair o objeto de interesse de cada frame da imagem. A partir das imagens segmentadas, o centróide da mão em cada frame é extraído e o movimento é representado por uma sequência de centróides. Neste mesmo repositório também existe o conjunto de dados relativo a ações físicas, o Vicon Physical Action, apresentado em [Theodoridis e Hu 2007], que usou um sistema avançado de captação de movimento chamado Vicon.

Dentro do escopo desse estudo, não foram encontradas iniciativas que oferecem uma sistematização automatizada para a construção de conjuntos de dados de re-

[1]Existem iniciativas que trabalham com segmentação de imagem ou reconhecimento de padrões manual, com o objetivo de fundamentar ou ilustrar alguma teoria de análise de gestos. Nestes casos, a amostra de análise é bastante restrita e as aferições são qualitativas e subjetivas. Trabalhos nessa linha são: [Chen et al. 2002], [Quek et al. 2002] e [Quek 2004]

[2]http://archive.ics.uci.edu/ml/index.html.

ferência no contexto de análise de gestos baseados em imagens. Mas foram analisados exemplos de iniciativas similares, pertencentes a diferentes contextos: a proposição da criação colaborativa de conjuntos de dados compostos por documentos Web, mediante a predefinição de "contexto" e "características" desejáveis para os dados, e a avaliação manual ou automática da aderência dos documentos adicionados [Barros et al. 2009]; o desenvolvimento de um sistema que constrói um conjunto de dados para suportar a tarefa de reconhecimento ótico de caracteres, considerando diferentes tamanhos e fontes diferentes e assumindo um fundo de imagem uniforme (branco) [Wahab et al. 2009]; o incentivo à padronização de conjuntos de dados de escrita à mão anotados, cobrindo diferentes estilos de escrita, como discutido em [Bhaskarabhatla et al. 2004].

Para melhor direcionar o leitor no entendimento da ferramenta desenvolvida, este artigo está organizado como segue: na Seção 2 é delimitado o escopo da ferramenta e são listados os requisitos funcionais elicitados para a composição desta sua primeira versão; as funcionalidades implementadas para atender aos requisitos elicitados são descritas na seção 3; as extensões já previstas para esta ferramenta são comentadas na Seção 4 juntamente com as considerações finais.

2. Definição de Escopo e Requisitos Funcionais

Um conjunto de dados de referência único para a área de análise de gestos não é possível e não faz sentido, visto as diferentes áreas de aplicação e a infinidade de possibilidades de gestos que podem ser analisados em cada uma delas. Além disso, as diferentes formas de resolver um problema de segmentação baseada em cor ou um problema de reconhecimento de padrões, também exige que os conjuntos de dados sejam construídos das mais variadas formas. Entretanto, é possível delimitar escopos para criação de conjuntos de dados que possam cobrir a avaliação de características de uma aplicação que realiza determinados tipos de análise usando determinadas abordagens de resolução de problemas.

Uma vez que pretende-se, com a ferramenta aqui descrita, atuar na área de análise de gestos baseada em imagens, a primeira delimitação de escopo se refere ao tipo dos dados que comporão o conjunto: **vídeos ou frames de vídeo** contendo o gesto estático ou dinâmico a ser analisado; ou **vetores de características** extraídas desses vídeos ou frames de vídeo. Assim, os primeiros requisitos funcionais elicitados são referentes à *aquisição de vídeos* e *transformação do vídeo em sequências de frames*. Também por se tratar de um sistema que atuará com cores, *converter os frames de vídeo para diferentes sistemas de cores* [Gonzalez e Woods 2000] é útil para verificar a sensibilidade das técnicas de segmentação de imagem em relação a discriminação da informação sobre cor. Os sistemas de cores escolhidos para uso são: RGB (sistema com informação sobre intensidade de cor vermelha, verde e azul), YCbCr (sistema com informação de luminância e cromaticidade), HSV (sistema com informação sobre a cor, saturação da cor e intensidade da cor) e escala de cinza (amostragem do espaço de cores).

Sistemas que atuam neste contexto de análise necessitam, geralmente, de uma fase de segmentação de imagem e/ou extração de características do gesto nela contido, antes que o reconhecimento do padrão propriamento dito possa ser realizado. Então, o segundo delimitador de escopo estabelecido aqui é: trabalhar com **segmentação de objeto de interesse**. Esta segmentação pode ser realizada sob diferentes abordagens, sendo que uma delas é usar um algoritmo de reconhecimento de padrões de cor e, isto posto, estabelece-

se mais um dos requisitos funcionais da ferramenta: criar os dados que servirão como entrada para os reconhecedores de padrões que segmentarão as imagens. A opção adotada na ferramenta em discussão é a organização desses dados como *vetores de características* obtidos a partir do mapeamento[3] de *recortes da imagem*, de forma que pedaços de imagens possam ser usados como exemplos de objeto de interesse para treinamento de algoritmos de reconhecimento de padrões. Outras estratégias de segmentação de imagens, que exijam outros formato de dados estão, por enquanto, fora do escopo desta ferramenta. O trabalho com extração de características está fora do escopo desta versão da ferramenta, contudo, a forma como a organização de dados está projetada, a inserção de funcionalidades nesta área é possível e está prevista nas intenções de extensões da ferramenta. Já a questão de **reconhecimento de padrões** é prevista, nesta versão, na funcionalidade de criação de conjuntos de dados de referência, configurando-se como um delimitador de escopo, como discutido mais à frente neste texto.

Os requisitos já citados estabelecem funcionalidades da ferramenta que geram diversos artefatos (vídeos, frames de vídeo, recorte de imagens e vetores de características). Esses artefatos comporão os conjuntos de dados de referência que, para atender as necessidades de aferição de desempenho e testes de sistemas, precisam estar **contextualizados e organizados**. Assim, requisitos funcionais para organização e descrição destes dados são atendidos pela ferramenta através da criação de *perfis de dados* e da *organização dos dados em diretórios* de acordo com os perfis sob os quais são obtidos. As informações consideradas nos perfis são: contexto ou tema (considerando tipo de fundo da imagem e características da cor do objeto de interesse), executor do gesto, tipo de gesto, sequência (informando se o vídeo possui um ou mais gestos ou partes/fases de um gesto), data de aquisição, espaço de cores usado no vídeo, tamanho das janelas de mapeamento de recortes de imagem em vetores de características.

Por fim, o último requisito funcional elicitado é a *criação do conjunto de dados*, que caracteriza a principal motivação para construção desta ferramenta. Os conjuntos são criados a partir da combinação de características que se deseja analisar em um sistema de análise de gestos (dentro do escopo supradescrito) e do tipo de teste que se deseja fazer. A ferramenta disponibiliza um ambiente gráfico onde algumas escolhas podem ser feitas pelo usuário. A partir desta escolha, o conjunto de dados em si é gerado e armazenado juntamente com um arquivo de descrição das características do mesmo.

A definição do escopo de atuação da ferramenta, bem como os requisitos elicitados são resumidos na Tabela 1. Neste resumo também observa-se o relacionamento entre as delimitações de escopo estabelecidas e as funcionalidades que as atendem.

3. Descrição das Funcionalidades

A ferramenta aqui discutida disponibiliza ao usuário funcionalidades que facilitam a aquisição, organização e descrição de dados persistidos em diretórios no sistema operacional do computador. As funcionalidades são graficamente apresentadas na Figura 1 sob a especificação do modelo de processo correlato, e são descritas nas seções subsquentes.

O modelo de interação com a ferramenta prevê a criação de uma sessão com identificação do perfil de aquisição de vídeos para que seja possível determinar como

[3]O mapeamento aplicado nesta ferramenta faz uma varredura nas imagens, ou nos recortes de imagem, usando janelas de pixels de diferente tamanhos (3x3, 5x5 ou 7x7 pixels).

Tabela 1. Definição de escopo e requisitos funcionais.

Definição de Escopo	**Requisitos Funcionais**	N^o
Contextualização e organização dos dados em diretórios	⋆ criação de perfis de dados ⋆ criação dos diretórios	1 2
Tratamento de vídeos	⋆ aquisição de vídeos ⋆ criação de frames de vídeo ⋆ conversão de frames (sistemas de cores)	3 4 5
Tratamento de janelas	⋆ criação de recortes da imagem ⋆ mapeamento dos recortes para vetores de características através da varredura com janelas de pixels	6 7
Criação do conjunto de dados	⋆ escolha e organização dos dados em conjuntos de referência para testes em segmentação de imagens e em reconhecimento de padrões	8

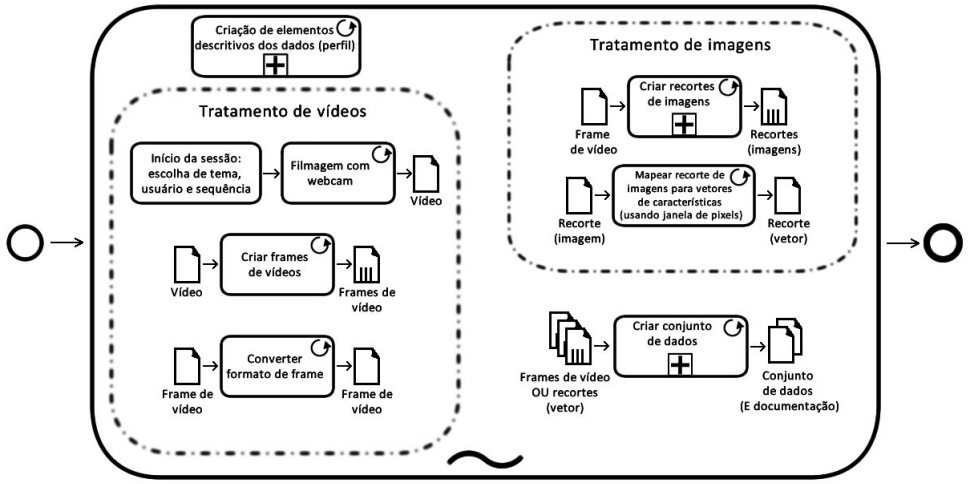


Figura 1. Modelo de processo: visão geral das ações permitidas na ferramenta.

organizar os vídeos nos diretórios de persistência de dados. Com exceção desta sequência exigida no processo, as outras funcionalidades ficam disponíveis para uso sob demanda, e os artefatos gerados na ferramenta são persistidos no sistema operacional do computador.

3.1. Criação de Perfis de Dados e Estrutura de Diretórios

As funcionalidades de criação de perfis de dados seguem o modelo padrão de sistemas de propósito geral. A diferença aqui é o tipo de informação que compõe os perfis e o uso destas informações para criar a estrutura de diretórios onde os artefatos (videos, imagens, vetores) serão armazenados. A hierarquia apresentada na Figura 2 mostra as informações usadas para organizar os dados e que poderão formar os perfis dos conjuntos a serem criados pelo usuário. Na Figura 3 são mostradas as telas que implementam a interação do usuário com as funcionalidades de criação de elementos que comporão a descrição dos perfis dos conjuntos de dados. Tais elementos representam a contextualização da informação persistida[4].

[4]A criação de usuário prevê apenas o cadastro de um identificador e por esta razão não foi ilustrada aqui.

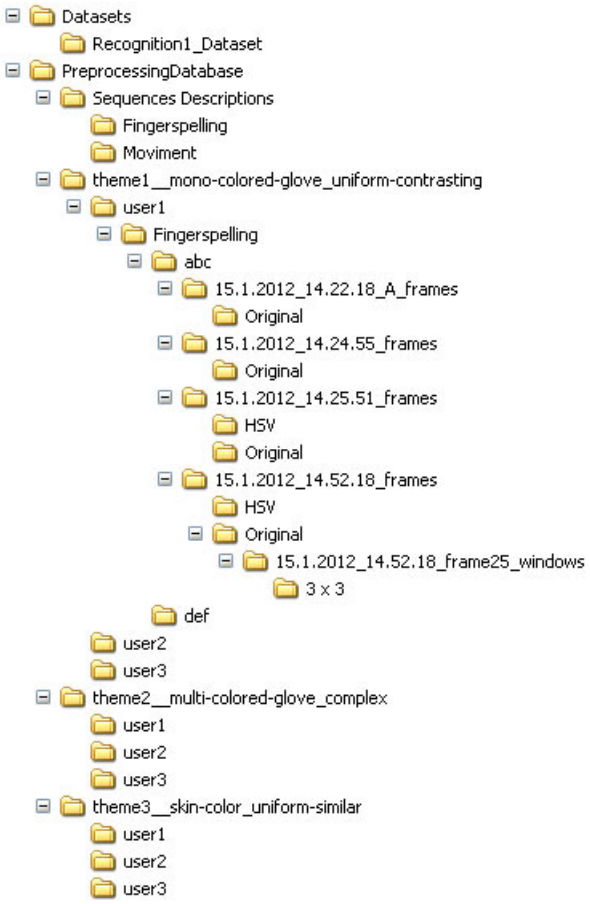


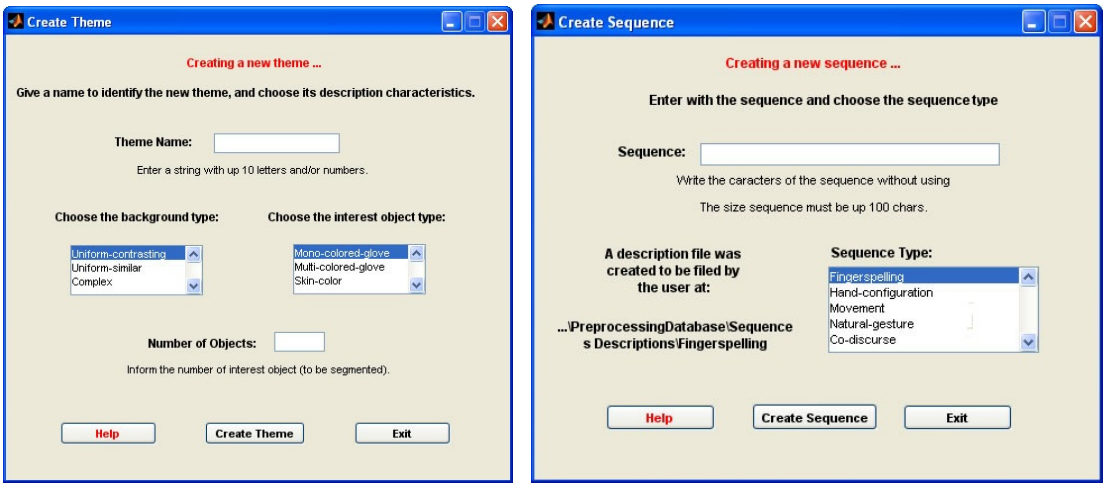
Figura 2. Exemplo de uma estrutura de diretórios criada a partir da interação com a ferramenta.

Na tela de criação de temas existem alguns atributos cujos valores precisam ser determinados pelo usuário. O atributo *background type* diz respeito ao tipo de fundo presente na imagem. As possibilidades de valores para este atributo (*uniform contrasting*; *uniform similar*; *complex*) informam a complexidade da tarefa de segmentação do objeto de interesse. Fundos uniformes contrastantes apresentam pouca variabilidade de cor de fundo e a cor predominante é bem diferente da cor do objeto de interesse. Fundos uniformes similares diferem do primeiro tipo por apresentarem a cor predominante parecida com a cor do objeto de interesse. Fundos complexos são aqueles que possuem uma grande variabilidade de cores, incluindo cores que estão presentes no objeto de interesse. Já o atributo *interest object type* deve ser escolhido de acordo com o uso, ou não, de luvas (de uma ou mais cores) pelo executor do gesto. Também é necessário informar o número de objetos a serem segmentados. No caso de ausência do uso de luvas ou no caso de uso de luvas de uma única cor, é esperada a existência de apenas um objeto de interesse[5]. Porém, no caso do uso de luvas coloridas é esperado que cada uma das cores represente um objeto de interesse diferente. Como exemplo considere uma luva colorida onde cada um dos dedos da luva é de uma cor diferente. Este esquema de luva colorida é útil para facilitar a descrição de gestos internos da mão, muito comuns em situações onde se quer analisar soletração em línguas de sinais (*fingerspelling*).

[5]Exceções podem ser admitidas conforme desejo do usuário.

Já na tela de criação de sequências é preciso informar o tipo da sequência de gestos que será realizada, e inserir informação descritiva sobre a sequência em um arquivo texto. Uma sequência significa que serão realizados um ou vários gestos de um determinado tipo. Os diferentes tipos de sequências atualmente previstos na ferramenta são: soletração (*fingerspelling*), configurações de mão (*hand-configuration*), movimentos (*Movement*), gestos naturais (*Natural-gesture*) e gestos que acompanham o discurso (*co-Discurse*)[6].

Para o correto armazenamento e descrição das informações sobre a sequência de gestos executada é necessário fornecer um nome para rotular cada um dos gestos realizados. Por exemplo, em uma sequência de gestos com dois movimentos, um na horizontal e outro na vertical, um nome possível para ela seria “AB”; no arquivo de descrição é esperado que o usuário indique que “A” significa “movimento horizontal” e “B” significa “movimento vertical”. Esse arquivo de descrição acompanhará o conjunto de dados quando a sequência correlata for escolhida para compor os dados do conjunto.



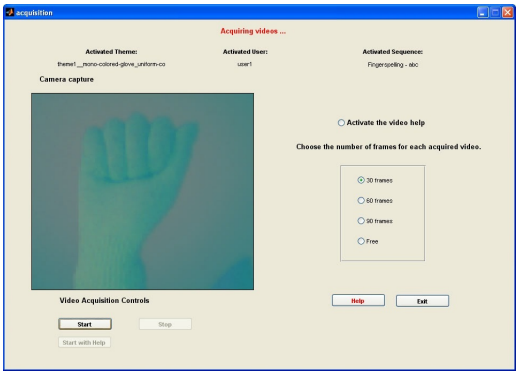
(a) Criação de novo tema (b) Criação de nova sequencia

Figura 3. Interface gráfica para criação de novos temas e sequências de gestos.

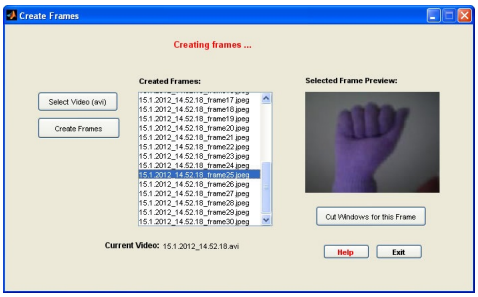
3.2. Aquisição de vídeo, criação de frames e conversão de vídeo

A Figura 4 mostra três telas de interação com o módulo de tratamento de vídeos (aquisição de vídeo, criação de frames e conversão de frames). Antes da interação com a funcionalidade de aquisição de um vídeo o usuário passou por uma tela de criação de sessão, onde descreveu o perfil do gesto que será captado nos vídeos. A informação sobre o perfil escolhido fica disponível na tela de aquisição de vídeo juntamente com algumas outras opções: visualização de *imagem de help*, disponível apenas para as sequências referentes à soletração e configuração de mão, e ativo apenas se usuário desejar; opção de *tempo de captação de vídeo* (em frames), com as opções de 30 frames, 60 frames, 90 frames ou livre (*free*). Neste último caso, a interrupção da captação do vídeo deve ser feita por meio do acionamento de um botão na interface gráfica. Nos casos anteriores, a interrupção é controlada pela ferramenta. A persistência dos vídeos bem como sua denominação é feita considerando a data e hora da aquisição e o rótulo a ele associado. A associação do rótulo é feita a partir do processamento do nome da sequência escolhida na sessão de aquisição.

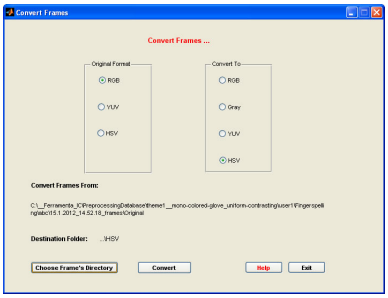
[6] A fronteira que separa os gestos naturais dos gestos que acompanham o discurso não é bem definida e depende no contexto de análise objetivado.



(a) Capturando vídeos



(b) Criando frames



(c) Convertendo frames

Figura 4. Interface gráfica para tratamento de vídeos.

Ainda no contexto de tratamento de vídeos o usuário pode escolher vídeos para que sejam gerados os frames correspondentes. Isso é necessário para que as demais funcionalidades da ferramenta possam operar criando os recortes de imagem e os vetores de características. E por fim, a conversão de frames para os diferentes sistemas de cores disponíveis na ferramenta pode ser feita também a partir da interação do usuário com a interface gráfica. Cada vídeo processado gera um diretório onde são armazenados seus frames, e cada conversão de sistema de cores solicitada gera um outro diretório de mesmo nível nomeado de acordo com o sistema de cores escolhido.

A partir da interface gráfica para a funcionalidade de criação de frames, o usuário tem a opção de seguir para as ações de gerenciamento de janelas. Se esta opção é escolhida, o usuário pode escolher um frame e proceder com as ações de recorte de imagem e mapeamento para vetor de características, discutidas na próxima seção.

3.3. Recorte de imagens e mapeamento para vetor de características

O recorte de imagens pode ser feito de duas maneiras: a) o usuário indica o tamanho em pixels que deseja para o recorte e marca posição do canto superior esquerdo do recorte com um clique de mouse; b) o usuário escolhe a opção de tamanho livre e marca o canto superior esquerdo e o canto inferior direito do recorte, com dois cliques de mouse, respectivamente. A ferramenta disponibilizará um visualização do recorte solicitado e, a partir da decisão do usuário, o armazenará. É necessário que o usuário informe se o recorte diz respeito a uma área do objeto de interesse (e qual objeto no caso de existirem vários na imagem) ou se trata-se de uma área do fundo da imagem. Essa indicação é necessária para a rotulação do recorte realizado. Veja na Figura 5 o ambiente onde o recorte é realizado.

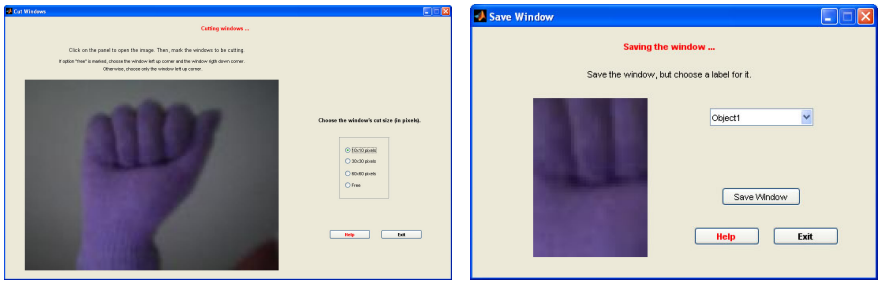


Figura 5. Interface gráfica para recorte de imagem.

Outra funcionalidade deste contexto é o mapeamento dos recortes de imagens para vetores de características. Como já mencionado neste artigo, esse mapeamento é feito através da varredura da imagem usando uma janela de pixels. O tamanho da janela de pixels deve ser escolhido pelo usuário, dentre as opções dadas pela ferramenta. Tanto as imagens recortadas quanto os vetores de características são persistidos pela ferramenta. Os recortes ficam associados aos frames dos quais são originários, e os vetores são armazenados em arquivos texto - um arquivo texto para cada recorte de imagem.

3.4. Escolha e organização dos dados em conjuntos de referência

Para a criação dos conjuntos de dados de referência, os usuários devem escolher as opções que representam as características que desejam aferir em seus sistemas de segmentação de imagens ou em seus sistemas de reconhecimento de padrões. As funcionalidades que implementam a criação do conjunto de dados possuem a mesma lógica de programação, diferindo apenas nas possibilidades de escolha de características que fornecem ao usuário.

Basicamente o usuário precisa escolher os diretórios que se referem às características que ele deseja inserir no conjunto de dados de acordo com algumas restrições que a ferramenta implementa. Além disso, o usuário pode escolher o tamanho de seu conjunto de dados ou deixar que a ferramenta construa o maior conjunto possível mediante os dados disponíveis dentro da combinação de características escolhida. Se o usuário determinar o tamanho do conjunto de dados, a ferramenta criará um subconjunto de dados com escolha aleatória dos elementos que o comporão (com base do conjunto completo disponível considerando as opções de características escolhidas pelo usuário).

O usuário fornece um nome para o seu conjunto e a ferramenta permite que ele escolha o local no sistema operacional onde o conjunto será gravado. Se o conjunto de dados se refere a vetores de características, ele é criado com a extensão *.data* acompanhado de um segundo arquivo de extensão *.names* com a descrição textual do seu conteúdo, baseado nas informações constantes nos nomes de diretórios, nomes de arquivos e conteúdo de arquivos textos constantes na organização da fonte de dados. Caso ele se refira a arquivos de vídeos ou imagens, ele é gravado em um diretório nomeado pelo usuário.

Na criação de conjuntos de dados para aferição de sistemas de segmentação de imagem, as opções de características são oferecidas da seguinte forma (veja a interface gráfica na Figura 6): escolha do objeto de interesse (apenas uma opção pode ser escolhida); a partir desta escolha o sistema verifica quais são os tipos de fundos disponíveis (o usuário pode escolher uma ou mais opções); considerando as combinação dos dois itens acima, a ferramenta mostra os temas disponíveis (o usuário pode escolher uma ou mais

opções); dado os temas escolhidos, é disponibilizado em série opções de usuários executores dos gestos (o usuário da ferramenta pode escolher uma ou mais opções); mediante as escolhas acima, um dos sistema de cores disponíveis e uma opção sobre o parâmetros de tamanho de janela de pixels disponíveis podem ser escohidos.

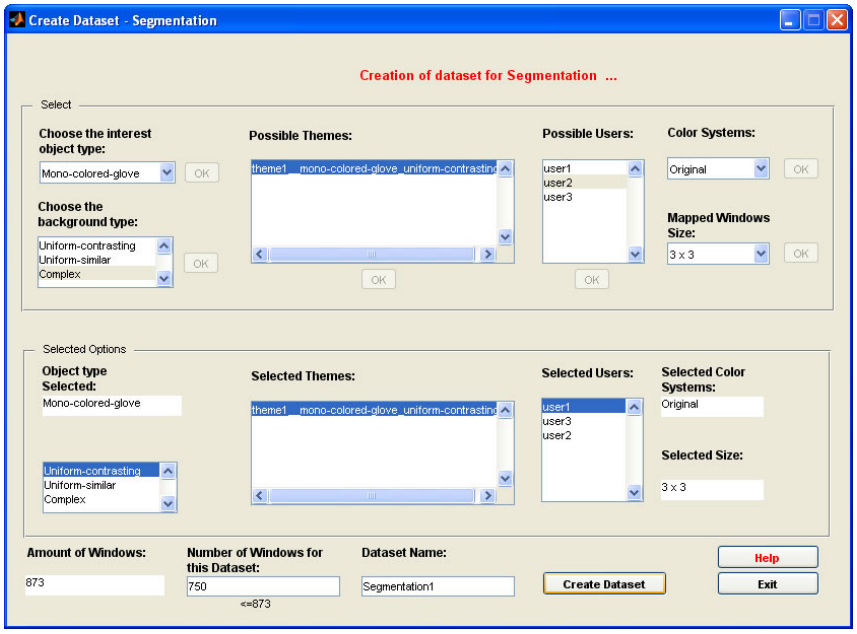


Figura 6. Interface gráfica para escolha de opções para criação de conjuntos de dados para testes com segmentação de imagens. Os objetos de interface são disponibilizados durante a interação: a cada escolha de característica, o objeto de interface que trata da próxima é disponibilizado. Na parte superior da janela o usuário realiza a interação. Suas escolhas são mostradas no centro da janela.

Já na criação de conjuntos para reconhecimento de padrões de gestos, a diferença está em algumas das opções de características oferecidas (veja Figura 7). As opções para este caso são as escolhas sobre: tipo do objeto de interesse (que neste caso podem ser um ou vários), tipo de fundo, temas, usuários executores de gestos, tipo de gestos (apenas um pode ser escolhido), tipo de frame (já com a realização da segmentação do objeto de interesse ou não) e sequências (várias podem ser escolhidas).

4. Considerações Finais

Este artigo apresentou uma ferramenta concebida para contribuir com a área de desenvolvimento de sistemas que realizam análise de gestos baseados em imagens, possibilitando a organização de dados que possam ser agrupados em conjuntos de dados de referência. O principal objetivo desta ferramenta é possibilitar que uma variedade de características interessantes na aferição de desempenho de tais sistemas sejam cobertas pelos conjuntos de dados usados nos testes dos mesmos. Desta forma espera-se contribuir para a eficiência e eficácia dos processos de testes comparativos, possibilitando uma análise adequada dos sistemas desenvolvidos.

Um requisito funcional importante que ainda está sendo implementado pela equipe de desenvolvimento da ferramenta é um módulo para extração de características que representem vetorialmente os diferentes tipos de gestos. Com essa funcionalidade,

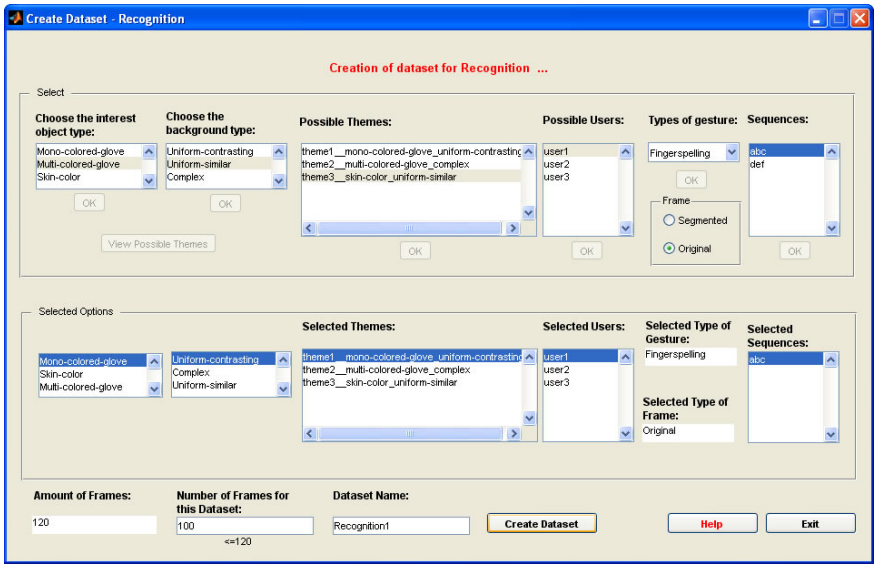


Figura 7. Interface gráfica para escolha de opções para criação de conjuntos de dados para testes com reconhecimento de padrões de gestos. A interação é similar à interface referente à criação de conjuntos para segmentação de imagens.

poder-se-á criar uma variedade maior de conjuntos de dados para teste de algoritmos de reconhecimento de padrões. Em relação a requisitos não funcionais, algumas melhorias ainda precisam ser feitas, como: melhorar o sistema de ajuda da ferramenta *help*, de forma a guiar a interação com o usuário, primando pelo uso correto da ferramenta e assegurando que a organização de diretórios seja usada da maneira mais adequada; possibilitar que os arquivos de descrição na funcionalidade de criação de sequência sejam feito de maneira tutorada (como um *wizard*).

Agradecimentos

Agradecemos ao Ministério da Educação (Programa de Educação Tutorial - PET) e à Fundação de Amparo à Pesquisa do Estado de São Paulo (processo 2011/04608-8).

Referências

Barros, R., Rodrigues Nt., J. A., Carneiro Filho, H. J. A., Ferreira, F. R. S., Fernandes, O. C., Silva, C. E. P., Ribeiro, A. L. G., Xexeo, G. B., e de Souza, J. M. (2009). A collaborative approach to building evaluated web pages datasets. In *Proceedings of the 2009 13th International Conference on Computer Supported Cooperative Work in Design*, pages 668–673, Washington, DC, USA. IEEE Computer Society.

Bhaskarabhatla, A., Madhvanath, S., Kumar, M., Balasubramanian, A., e Jawahar, C. (2004). Representation and annotation of online handwritten data. In *Frontiers in Handwriting Recognition, 2004. IWFHR-9 2004. Ninth International Workshop on*, pages 136–141.

Chen, L., Harper, M., e Quek, F. (2002). Gesture patterns during speech repairs. In *Proceedings. Fourth IEEE International Conference on Multimodal Interfaces*, pages 155–160. IEEE Comput. Soc.

Dias, D. B., Madeo, R. C. B., Rocha, T., Bíscaro, H. H., e Peres, S. M. (2009). Hand movement recognition for brazilian sign language: A study using distance-based neural networks. In *International Joint Conference on Neural Networks*, pages 697–704.

Duchowski, A. (2002). A breadth-first survey of eye-tracking applications. *Behavior Research Methods*, 34:455–470.

Ekman, P., F. W. e Hager, J. (2002). *Facial Action Coding System. A Human Face*. Salt Lake City, USA.

Gonzalez, R. C. e Woods, R. E. (2000). *Processamento de Imagens Digitais*. EDGAR BLUCHER.

Just, A., Rodriguez, Y., e Marcel, S. (2006). Hand posture classification and recognition using the modified census transform. In *7th International Conference on Automatic Face and Gesture Recognition*, pages 351–356, Southampton.

Kadous, M. W. (1995). Grasp: Recognition of australian sign language using instrumented gloves.

Kendon, A. (1980). *The Relation Between Verbal and Nonverbal Communication*, chapter Gesticulation and speech: Two Aspects of the Process of Utterance, pages 207–227.

Kita, S., van Gijn, I., e van der Hulst, H. (1998). Movement phases in signs and co-speech gestures, and their transcription by human coders. In Wachsmuth, I. e Frohlich, M., editors, *Gesture and Sign Language in Human-Computer Interaction*, volume 1371 of *Lecture Notes in Computer Science*, pages 23–35. Springer Berlin / Heidelberg.

Madeo, R. C. B., Peres, S. M., Bíscaro, H. H., Dias, D. B., e Boscarioli, C. (2010). A committee machine implementing the pattern recognition module for fingerspelling applications. In *Proceedings of Symposium on Applied Computing*, pages 954–958.

McNeill, D. (1992). *Hand and Mind*. University of Chicago Press.

Mitra, S. e Acharya, T. (2007). Gesture recognition: A survey. *IEEE Trans. on Systems, Man, and Cybernetics - Part C: Applications and Reviews*.

Pantic, M., Pentland, A., Nijholt, A., e Huang, T. (2007). Human computing and machine understanding of human behavior: a survey. volume 44512007 of *Lecture Notes in Artificial Intelligence*, pages 47–71, Berlim. Springer-Verlag.

Quek, F. (2004). The Catchment Feature Model: A Device for Multimodal Fusion and a Bridge between Signal and Sense. *EURASIP Journal on Advances in Signal Processing*, 2004(11):1619–1636.

Quek, F., McNeill, D., Bryll, R., Duncan, S., Ma, X., Kirbas, C., McCullough, K., e Ansari, R. (2002). Multimodal human discourse: gesture and speech. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 9(3):171–193.

Theodoridis, T. e Hu, H. (2007). Action classification of 3d human models using dynamic anns for mobile robot surveillance. In *Robotics and Biomimetics, 2007. ROBIO 2007. IEEE International Conference on*, pages 371 –376.

Wahab, M., Amin, H., e Ahmed, F. (2009). Shape analysis of pashto script and creation of image database for ocr. In *Emerging Technologies, 2009. ICET 2009. International Conference on*, pages 287 –290.