Interação de objetos virtuais através de posturas da mão

Pedro H. L. da Silva, Luis Rivera, Annabell del Real, Fermin Tang

Laboratório de Ciências Matemáticas – LCMAT
Universidade Estadual do Norte Fluminense - UENF
Av. Alberto Lamego, 2000; CEP 28015-620
Campos dos Goytacazes – RJ – Brazil

pedrolmota@gmail.com, {rivera, annabell, tang}@uenf.br

Abstract. *The means of interaction between user/computer are mostly through keyboards, mice, keypads, etc. Recently have been emerging alternative means of interaction with the computer, such as gestures that produce commands, using techniques of pattern recognition and computer vision. In this paper is developed a virtual environment where user interaction takes place through the recognition of hand defined postures. This system is built on the basis of segmentation algorithms so that the user's hand was captured through a webcam, characterized using moments of Hu and represented in a multilayer Perceptron neural network for classification. For the interaction of objects in the 3D virtual environment, positions are obtained from the changes of images of the hand catches the web.*

Resumo. *Os meios de interação entre usuário/computador são predominantemente através de teclados, mouse, keypads, etc. Recentemente vem surgindo meios alternativos de interação com o computador, como os gestos que produzem comandos, usando técnicas de reconhecimento de padrões e visão computacional. Neste trabalho é desenvolvido um ambiente virtual cuja interação com o usuário acontece através do reconhecimento de posturas de mão predefinidas. Esse sistema é construído sobre a base de algoritmos de segmentação de forma que a mão do usuário fosse capturada através de uma webcam, caracterizada usando momentos de Hu e representada em uma rede neural Perceptron multicamada para sua classificação. Para a interação dos objetos no ambiente virtual 3D, as posições são obtidas a partir das variações das imagens da mão capturas pela webcam.*

1. Introdução

Por muito tempo os meios de interação com o computador têm se limitado ao teclado, mouse, keypads, trackballs, etc. Esses dispositivos, apesar de adequados, limitam a naturalidade com a qual se interage com os computadores. Recentemente, com o desenvolvimento de computadores mais potentes e com o crescimento do interesse por meios mais envolventes com sistemas computacionais, vêm se desenvolvendo formas de interação com gestos entre o usuário e o computador. A interação com gestos utiliza técnicas de visão computacional combinadas com métodos de reconhecimento de padrões.

O uso de gestos na comunicação entre pessoas é comum, geralmente são usados implicitamente como complemento da fala. Geram-se gestos através de movimentos de partes do corpo, como as mãos, o rosto, os olhos e a cabeça ou o movimento do corpo inteiro. Os mudos, por exemplo, utilizam os gestos de mãos como principal forma de comunicação, e é, também, com as mãos que as pessoas manipulam os objetos em geral. Sendo assim, cada gesto traz consigo uma mensagem que pode ser obtida pelo reconhecimento visual das posturas que o compõem. Mandos gerados pela mensagem de gestos têm sido muito utilizados em jogos por

computador, os chamados jogos fisicamente interativos com princípios de imersão, popularizados ultimamente com os produtos da tecnologia Kinect.

Na pesquisa de uso de gestos na interação homem-computador, o problema de correto entendimento do gesto através do reconhecimento das posturas envolve diversos aspectos a se desenvolver, como, por exemplo, detectar as posturas a partir de um conjunto de frames de um vídeo e como representar essas posturas de forma eficiente; ou seja, como caracterizá-las, e por fim, como executar o reconhecimento dessas posturas. Para a solução de cada um desses aspectos existem técnicas e abordagens que devem ser selecionadas dependendo da aplicação específica.

No que diz respeito aos dispositivos de captura, existem trabalhos como [Ahn+11] [Gonçalves+12] [Mahbub+12] [Mitra+07] que usam categorias de dispositivos sofisticados de captura para reconhecimento de gestos. Por exemplo, em [Ahn+11] é usada uma câmera com sensor de movimento DVS-câmeras (*dynamic vision sensor câmeras*). Em [Gonçalves+12] e [Mahbub+12] são usados câmeras com sensor Kinect com tendências eficientes para reconhecimento de gestos aplicados a jogos e seguranças. Porém, essas câmeras não são de usos comuns como as simples webcams. Os dispositivos intrusivos, como marcas de rastreamento, luvas, roupas especiais, etc., são orientados para aplicações específicas não comuns. Mitra et al. [Mitra+07] analisam métodos de reconhecimento de gestos com diversas abordagens, incluindo dispositivos intrusivos. Neste caso, o interesse é a análise de sequência de imagens capturadas por uma câmera, tal como a webcam de uma laptop ou uma webcam tradicional, para operar em qualquer ambiente e lugar, sem precisar de outros dispositivos adicionais intrusivos.

O trabalho se organiza da seguinte forma: na Seção 2 abordam-se reconhecimentos de gestos e trabalhos relacionados. Na seção 3 se formula o modelo de reconhecimento de gestos para a interação no ambiente virtual; na Seção 4 apresentam-se os resultados do modelo proposto com exemplos de interação com ambientes virtual, e finalmente na Seção 5 conclui-se complementando com trabalhos futuros.

2. Gestos e posturas de mão

Um gesto da mão é composto por uma sequência de posturas de mão interligadas em um curto espaço de tempo. Por exemplo, o gesto de dar "tchau" para uma pessoa é a movimentação de um lado para outro da mão aberta. Diversos trabalhos foram desenvolvidos com propósitos similares utilizando diversas formas de reconhecimento de posturas por meio da visão computacional. Para isso, as imagens com as posturas de mão devem ser representadas com uma menor quantidade de informação, o que se conhece como caracterização, para serem então classificadas tendo suas posturas reconhecidas. Uma das formas utilizadas para caracterizar imagens estáticas é através do uso de invariantes. Também, existem diversos processos de classificação, sendo que, um dos mais utilizados são os classificadores usando redes neurais, como abordado por Junior et al [Junior+07].

Existem diversos trabalhos que utilizam gestos da mão para diversas aplicações, entre elas reconhecimento da linguagem de sinais, e para controle de objetos virtuais, de teleconferência, etc. Por exemplo, Chen et al. [Chen+07] criaram uma ferramenta para a detecção de quatro posturas de mão, a posição com dois dedos, a palma da mão, o pulso e a posição com o dedo mínimo em destaque usando características Haar-like baseado no algoritmo desenvolvido por Viola e Jones [Viola+01]. Berry e Pavlovic [Berry+98] integraram gestos de controle, usando uma máquina de estados finitos para modelar a dinâmica dos movimentos, no ambiente virtual Battle-Field. Bretzner et al. [Bretzner+02] criaram um sistema para reconhecimento de gestos da mão, onde os gestos são representados em termos de características de hierarquias de imagens em cores em multi-escala, posição e orientação. A

mão é representada por um modelo que consiste na representação da palma da mão e os cinco dedos. O reconhecimento do gesto é realizado através de métodos estatísticos. Elmezain et al. [Elmezain+09] desenvolveram um sistema para reconhecer gestos de caracteres do alfabeto (A-Z) e números (0-9) em tempo real usando modelos ocultos de Markov (Hidden Markov Models - HMM). O sistema foi desenvolvido em três estágios, segmentação automática e pré-processamento de regiões da mão, extração de características e classificação. Carneiro et al. [Carneiro+09] implementaram um sistema que reconhece as 26 letras do alfabeto da LIBRAS através de uma rede neural Perceptron de múltiplas camadas e de uma rede SOM (Self-Organizing Map) que realiza uma pré-classificação a partir dos momentos invariantes de Hu.

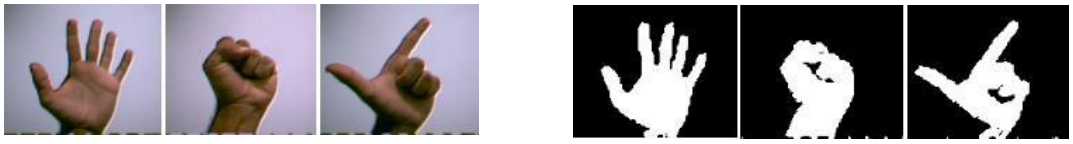
3. Modelo de reconhecimento de posturas

Neste trabalho é formulado um modelo de reconhecimento de posturas de mão para a manipulação de objetos como cubos e esferas em um ambiente virtual. O ambiente é composto por um piso e paredes delimitando a área do ambiente. No piso são dispostos os objetos geométricos a serem selecionados e movimentados para qualquer posição no ambiente. Para esse propósito, se estabelecem três posturas básicas: *movimentar*, *selecionar* e *segurar*. A postura de *movimentar* é representada, como mostra a Figura 1(a), pela mão aberta; a postura *selecionar* pela mão em forma de L e *segurar* pela mão fechada. O modelo proposto está estruturado em cinco módulos: captura e segmentação, caracterização, treino, classificação, e interação no ambiente virtual.

3.1. Captura e segmentação

Cada frame obtido pela câmera constitui uma imagem estática. Essa imagem é segmentada, onde se que decide quais os pixels que correspondem à cor da pele, e quais fazem parte do fundo. É considerada uma tonalidade de pele específica para extrair somente a mão das outras partes da imagem.

Existem várias formas e métricas para se obtiver o tom da pele. Segundo Vezhnevets et al. [Vezhnevets+03] entre os métodos mais comuns para a detecção de tom da pele estão: *Normalized lookup Table* (LUT) que usa uma tabela de valores e quantizados de tons de cores; *Classificador Bayes* que verifica que dado um ponto *p* de cor *c*, se esta cor tem uma frequência maior que um limite *k*, para pertencer à pele; e *Região da pele definida explicitamente*, onde os valores correspondentes ao tom de pele serem definidos explicitamente em um determinado espaço de cores. Neste trabalho, utiliza-se a região da pele definida explicitamente no espaço de cores YCbCr, para $77 \leq Cb \leq 127$ e $133 \leq Cr \leq 173$ definido por Mahmoud [Mahmoud08], sendo este o método mais eficiente que o equivalente no espaço de cores RGB.



(a) (b)

Figura 1: Posturas para a interação no ambiente virtual (a) e segmentação das imagens no espaço YCbCr (b).

As imagens segmentadas ainda são impuras, devido aos ruídos e segmentos adicionais não desejados, que dificultam a identificação plena do gesto. As operações morfológicas (erosão, dilatação, suavização) permitem limpar os ruídos e os *blobs* [Bradski08][Gonzales+10]. Um problema detectado nesta fase foi que o resultado da segmentação varia de acordo com a iluminação do ambiente. Porque como se trata de uma limiarização sobre os canais da imagem, a incidência de luz no ambiente varia os valores das cores em cada canal. Ambiente com

iluminação imprópria, como pouca iluminação ou com muita iluminação, afeta o resultado da segmentação.

3.2. Caracterização

Cada frame obtido pela câmera constitui uma imagem estática. Essa imagem é segmentada, onde se que decide quais os pixels que correspondem à cor da pele, e quais fazem parte do fundo. É considerada uma tonalidade de pele específica para extrair somente a mão das outras partes da imagem. Esse processo consiste em analisar a imagem pré-processada e buscar pelos padrões de interesse. Nesta fase, a partir da imagem, são obtidos valores numéricos que representam eficientemente a imagem. Esse conjunto de valores forma o vetor de características da imagem fornecida.

A seleção de boas características para o reconhecimento de posturas de mão é uma fase crucial no processo e pode determinar o sucesso ou falha do algoritmo em uso. O fato de que a mão humana variar grandemente nas posições que pode assumir e movimentos que pode realizar é a principal razão pela qual a escolha de um bom conjunto de características é importante. Esse conjunto de características deve idealmente descrever a postura e/ou gesto de forma única, de forma que cada diferente posição da mão deve prover um conjunto de boas e diferentes características para um reconhecimento confiável.

O reconhecimento de posturas da mão é possível através da extração de características como direção da mão e dedos, ponta dos dedos, contorno da mão, em resumo características geométricas em geral. Existem também muitos trabalhos que se utilizam do fato de que a mão humana ter aproximadamente o mesmo matiz, saturação e variarem em seu brilho. Em outras palavras, sendo as mãos são detectadas pelo tom da pele. Essa é uma abordagem simples e muito utilizada para detectar a mão a partir de uma imagem. Elmezain et al. [Elmezain09] utilizam a localização, orientação e velocidade da trajetória da mão. Os momentos invariantes de Hu [Hu62] são extratores de imagens bastante eficientes apenas para um conjunto de sete parâmetros insensíveis às deformações rígidas, como translação, rotação, escala e espelhamento. O momento de invariantes de Hu tem sido amplamente utilizado em diferentes aplicações desde sua publicação [Wu+99]. Neste trabalho, usa-se invariante de Hu para caracterizar cada frame segmentado em um vetor de sete elementos. Os sete parâmetros, em detalhe em [Gonzales+10], são:

$$\phi_1 = \eta_{20} + \eta_{02}; \quad \phi_2 = (\eta_{20} + \eta_{02})^2 + 4_{11}^2; \quad \phi_3 = (\eta_{30} - 3\eta_{12})^2 + (3\eta_{21} - \eta_{03})^2$$

$$\phi_4 = (\eta_{30} - \eta_{12})^2 + (\eta_{21} - \eta_{03})^2$$

$$\phi_5 = (\eta_{30} - 3\eta_{12})(\eta_{30} + \eta_{12})\left[(\eta_{30} - \eta_{12})^2 - 3(\eta_{21} + \eta_{03})^2\right] + (3\eta_{21} - \eta_{03})(\eta_{21} + \eta_{03})\left[3(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2\right]$$

$$\phi_6 = (\eta_{20} - \eta_{02})^2\left[(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2\right] + 4\eta_{11}(\eta_{30} - \eta_{12})(\eta_{21} + \eta_{03})$$

$$\phi_7 = (3\eta_{21} - \eta_{30})(\eta_{30} + \eta_{12})\left[(\eta_{30} + \eta_{12})^2 - 3(\eta_{21} + \eta_{03})^2\right] + (3\eta_{21} - \eta_{03})(\eta_{21} + \eta_{03})\left[3(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2\right]$$

Os sete momentos de Hu, calculados através da biblioteca OpenCV que implementa as expressões prévias, para as três imagens da Figura 1(b) são apresentadas na Tabela 1. Observe que os valores são tão pequenos, em particular ϕ_5 e ϕ_7, praticamente zeros que fazem a rede neural, Devido ao fato que os valores de Hu são pequenos e dispersos, estes devem ser normalizados em função da media e da variância do conjunto dos valores de m posturas do treino. O valor normalizado N_{in} do valor do momento ϕ_n da postura i (ou seja, ϕ_{in}), para

$n = 1,...,7$ e $i = 1,...,m$ é dado como $N_{in} = \frac{\phi_{in} - \bar{\phi}_n}{\sigma_n}$, onde $\bar{\phi}_n$ e σ_n são, respectivamente, é a media aritmética e desvio padrão dos valores ϕ_n das m posturas do treino. Para o processo de treino é usado um número m grande de posturas, por tanto, também os valores normalizados variam dos exemplos apresentados.

Tabela 1: Valores pequenos dos momentos de três imagens.

Posturas	Mão aberta	Mão fechada	Mão L
ϕ_1	0.0007925	0.0006592	0.0009011
ϕ_2	3.4789e-08	3.4504e-08	5.315e-08
ϕ_3	1.3930e-11	5.1841e-13	2.5059e-10
ϕ_4	9.6427e-12	9.7159e-15	4.9476e-11
ϕ_5	1.9598e-23	5.8163e-28	-1.1871e-21
ϕ_6	1.0006e-15	1.6137e-18	-3.1229e-15
ϕ_7	1.1002e-22	-3.703e-28	5.3796e-21

3.3. Treino e classificação

Segundo Mitra e Acharya [Mitra+07], os métodos mais comuns em reconhecimento de gestos são: HMM (Hidden Markov Model), máquinas de estados finitos, redes neurais e outros como filtros de partículas, etc.. HMMs são usados em [Mitra+07], [Yoon+01], [Chen+03] [Mohandes+12]. Máquinas finitas por Manresa et al. [Manresa+00], Wu e Huang [Wu+99]. A rede neural, em particular Perceptron, é utilizada em vários trabalhos de reconhecimento de gestos de mão como em [Murthy+10] e [Symeonidis00].

Neste trabalho, para o reconhecimento de posturas é utilizada uma rede neural Perceptron de três camadas, tal como ilustra a Figura 2, com sete neurônios de entrada, devido a que se têm sete valores normalizados de momentos para cada imagem de entrada. São considerados três neurônios de saída, já que se têm três gestos a se identificar (mão aberta, fechada e em L). A Camada oculta é composta por seis neurônios, cuja determinação é feita através de várias tentativas de modelagem a rede, com diferentes números por vez. Observou-se que com seis neurônios de camada oculta optem-se melhor taxa de acertos no reconhecimento.

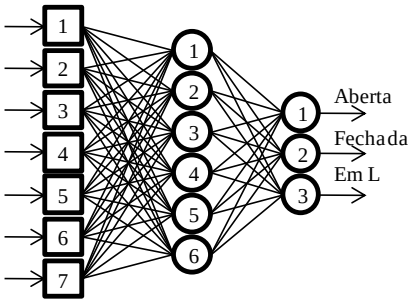


Figura 2: Rede neural Perceptron de três camadas.

Para o treino foram 110 imagens das três posturas (42 de mão aberta, 33 de mão fechada e 35 de mão em L), em diferentes formas (escala, orientação e posição) e com a colaboração de varias pessoas.

3.3. Treino e classificação

O ambiente criado usando Ogre3D consiste de piso (plano z=0), paredes (planos laterais x=-xc e x=xc, e plano de fundo y = yc, para xc e yc constantes), e outros objetos, com as características de realismo em 3D. Os objetos a serem movimentados no ambiente virtual são cubos e esferas. A movimentação dos objetos se realiza variando a posição das coordenadas x, y e z do centroide do boxe envolvente do objeto. Por tanto coordenadas do espaço virtual é dado por (x, y, z).

Neste cenário existe o problema de mover o indicador do mouse no ambiente 3D quando tipicamente este se movimenta no plano 2D da janela de projeção da cena virtual, sendo assim poderia se confundir movimento no eixo z com o do eixo y. Para resolver esse problema se considera a relação do espaço virtual com o espaço real. Chame-se espaço real ao espaço 3D, representado pelas coordenadas (X, Y, Z), frente do monitor onde se movimenta o braço do usuário para gerar as posturas, e de onde a câmera captura as imagens 30 frames por segundo acionado por OpenCV. Esse espaço é definido como sendo o eixo focal da câmera, dado por eixo Z, passando pelo centro geométrico das imagens capturadas, por tanto pelos centros geométricos dos objetos no ambiente real, neste caso a mão.

A relação entre esses dois espaços é estabelecida considerando o plano horizontal inferior (piso) do ambiente virtual como sendo o plano imaginário vertical perpendicular ao eixo focal da câmera (ver a Figura 3). Nesse plano imaginário (um dos planos) variam as coordenadas (X, Y). Um plano imaginário deve estar a uma distância inicial onde acontecerão as primeiras capturas das imagens da mão em movimento. Assim, quando a mão se movimenta nesse plano, por exemplo, sem perder a generalidade, como se fosse paralelo à tela no monitor, no vertical (eixo Y) será movimento de profundidade (eixo y) no piso do ambiente virtual, e horizontal (eixo X) do monitor será o eixo X do piso virtual. A analogia de tela do monitor com o plano vertical real imaginário é só um caso particular quando o eixo focal da câmera é considerado como sendo perpendicular ao plano monitor, porque a câmera pode estar focando para qualquer espaço do ambiente real.

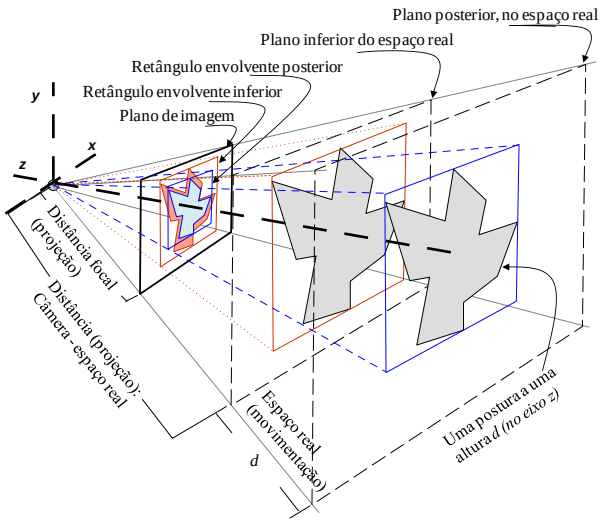


Figura 3: Esquema de relação entre os espaços.

A altura no ambiente virtual (eixo z), perpendicular ao piso, é relacionada com eixo focal da câmera que é considerada com eixo Z do ambiente real. Observa-se que quando a mão se afasta do plano inicial (ou se afasta mais da câmera) a imagem da mão fica cada vez menor. Essa diferença dos tamanhos das imagens das mãos em cada frame permite estabelecer a variação de posição no eixo Z do ambiente virtual. A unidade de variação no eixo Z é uma

variação de $\Delta z = 5$ unidades do tamanho da mão nas imagens de frames consecutivos. Esse valor é estimado através de várias experiências de movimentação da mão e indicador de mouse no ambiente virtual. Por cada frame atual existirá um frame antigo. Se não houver um frame antigo, será considerada a ocorrência do plano vertical imaginário, portanto se define o primeiro frame. Nos passos seguintes existirão os frames atual e antigo. Para computar as diferenças de tamanhos dos objetos, deve se computador *blobs* (caixas ajustadas retangulares) envolvendo, respectivamente, ao atual e ao antigo (que já tem dos passos anteriores). A diferença dos tamanhos dos dois objetos é a diferença das áreas de seus respectivos *blobs*. Este processo funciona para objetos da mesma postura (previamente reconhecida), caso contrário se considera como o frame ocorrendo por primeira vez.

A movimentação do indicador do mouse se é ativado quando é detectado movimento das mãos com posturas *aberta* e *fechada*. A *mão em L* em movimento não ativa o variação de posição do indicador do mouse. Isto porque a mão aberta indica só movimentação no ambiente em busca de um objeto, enquanto a mão em L indica a seleção do objeto encontrado. Na transição de aberta para L a mão pode se mover sem desejar, enquanto o indicador do mouse está bem próximo do objeto desejado tem possibilidade de selecionar o objeto. A proximidade é determinada pela relação entre distância entre o centro da mínima esfera envolvente do objeto é a posição do indicador do mouse e o raio dessa esfera. Uma vez o objeto desejado for selecionado, a postura de L deve passar para postura fechada, como indicando que o objeto foi pego, para poder levar o objeto a qualquer posição do ambiente virtual. As sequências das posturas devem ser respeitadas, tal como especificada pela máquina de estados da Figura 4, onde de mão aberta não pode passar para mão fechada, de mão fechada não pode passar para mão em L.

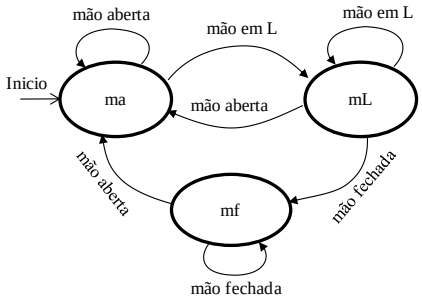


Figura 4: Máquina de estados para sequência de posturas.

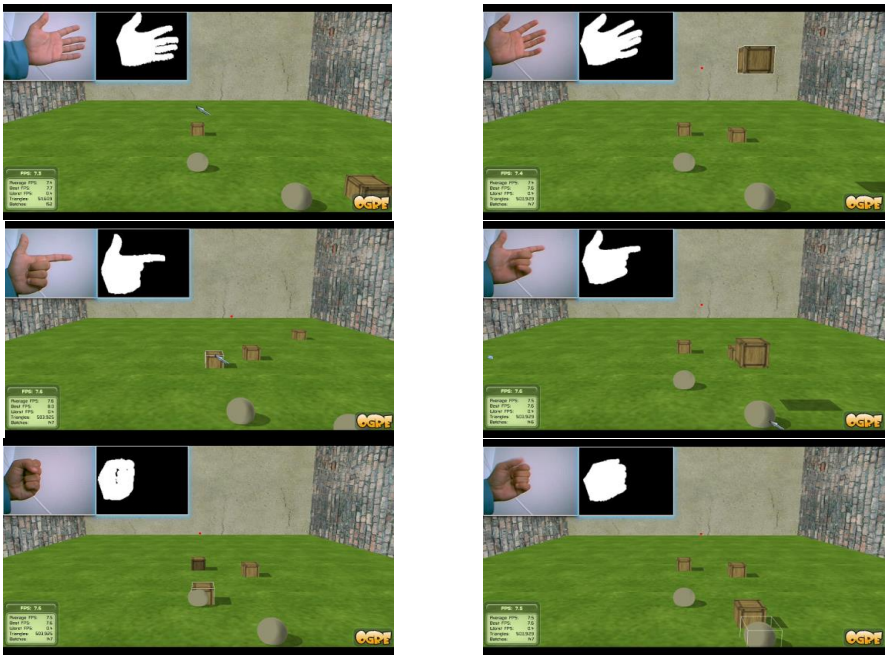
4. Interação com objetos virtuais

No inicio todos os elementos do ambiente, como ponteiro de mouse, câmera de captura, etc., são inicializados com valores defaults. Nos instantes de execução, qualquer postura que não seja similar às três definidas é ignorada, permanecendo no último estado válido. Tal como ilustrada pela Figura 4, o inicio de uma iteração deve estar no estado de mão aberta (*ma*), sendo possível identificar mão-aberta nos seguintes quadros, por tanto permanecer no mesmo estado *ma*. Cada variação de posição de mão aberta deve permitir a variação da posição do ponteiro do mouse no ambiente virtual. Veja que a única postura válida permitida para sair do estado *ma* é mão em L, que permite passar para o estado *mL*. Nesse estado deve ser verificada a proximidade do ponteiro do mouse com cada objeto movível. Se existir algum objeto satisfazendo a condição de proximidade, então o objeto é selecionado mostrando o bounding box do objeto, caso contrario o controle deve voltar para o estado *ma* com postura de mão aberta. A transição do estado *mL* para *mf* (mão fechada) só acontece quando algum objeto for selecionado no estado *mL* e o gesto *mf* for identificado, permitindo que o objeto selecionado seja movimentado pelo ambiente virtual 3D, de acordo o movimento da mão fechada no espaço

real. As três imagens que mostra a Figura 5(a) ilustram um ciclo completo da iteração com objeto virtual: buscar, selecionar e movimentar.

Como ocorre no mundo real, algumas vezes uma pessoa pega um objeto, levanta para analisar e solta se esse objeto não for de seu agrado, e selecionado outro objeto e pega. O sistema desenvolvido permite realizar também essa ação, onde um objeto que está sendo movimentado e é solto, digamos a uma altura *d*, e cai até se posicionar no piso, enquanto se busca outro objeto, se seleciona e se movimenta. A Figura 5(b) ilustra essa sequência de ações.

Tenha em conta que neste trabalho não foi incorporado os efeitos físico de colisões nem detecção de choques por consideramos esses tópicos fora dos objetivos de deste trabalho que é reconhecimento de gestos para interação com objetos virtuais. Neste sistema observou-se, em condições de ambiente real com uma iluminação apropriada, uma porcentagem alta (87% dos casos) de acertos de reconhecimentos dos três gestos. Em condições adversas, em ambiente de muita ou pouca luz, a taxa de acertos diminui para aproximadamente a 60% dos casos, já seja desconhecendo o gesto ou confundido com outros gestos.



(a) (b)

Figura 5: Gestos como comandos de interação com objetos no ambiente virtual.

5. Conclusões e trabalhos futuros

O reconhecimento de gesto geralmente se realiza como reconhecimentos de posturas. No reconhecimento de uma postura deve ser considerado um conjunto de situações que parte do corpo humano, neste caso a mão, pode estar representando ou tentando representar essa postura, que a câmera captura, em melhor dos casos como um gesto aproximado a um suposto gesto ideal. Quando se trata de captura de imagem por uma câmera web existem casos de variações de forma, de tonalidade, de influência da luz, sombra, dos objetos do ambiente onde acontecem as capturas. Uma correta segmentação das imagens pode colocar em evidência o gesto para uma melhor caracterização. Neste trabalho se realizam reconhecimentos de três gestos bastante diferenciados, mas ainda assim ocorrem variações de forma que foram consideradas redundantes ou não reconhecíveis. Mas felizmente os erros de variações em condições normais foram poucas, que comparadas com trabalhos similares estão dentro das margens de erros considerados.

A rede neural Perceptron, implementado neste trabalho, com uma data de entrada preparada apropriadamente, como o caso de momentos invariantes de Hu normalizados, permitiram uma representação eficiente de conhecimento dos gestos para uma eficiente classificação no processo de reconhecimento. Os valores classificados, complementados com as operações de posicionamento dos objetos no ambiente 3D, permitiram realizar um critério de imersão num ambiente virtual para poder movimentar os objetos virtuais unicamente realizando gesto de mão tal como estivesse manipulando os objetos no espaço real.

Nessa abordagem ainda restam muitos trabalhos para realizar, como o de diminuir o número de erros de acerto no reconhecimento, considerar mais posturas para identificar gestos completos, considerar outros paradigmas de representação de conhecimentos. Adicionar os processos físicos para gerar os choques, e por tanto também, considerar os método de detecção de colisões entre objetos, etc.

Referências bibliográficas

[Ahn+11] Ahn, E.Y.; Lee, J. H.; Mullen, T.; Yen, J. Dynamic Sensor Camera Based Bare Hand Gesture Recognition. IEEE Proceedings of Symposium on Computational Intelligence for Multimedia, Signal and Vision Processing (CIMSIVP), 2011. Pags. 52-59.

[Al-Rounsan+09] Al-Rousan, M.; Assaleh, K.; e Tala'a, A. Video-based signer-independent Arabic sign language recognition using hidden Markov models. Applied Soft Computing, 9, (2009), 990-999.

[Bradski08] Bradski, G. Kaebler, A. Learning OpenCV: O'reilly, 2008, 580 ps.

[Berry+98] Berry, G. Small-wall: A multimodal human computer intelligent interaction test bed with application. Master Dessertation, University of Illinois at Urbana-Champaign, 1998.

[Bretzner+02] Bretzner, L.; Laptev, I.; Lindeberg, T. Hand gesture recognition using multi-scale colour features, hierarchical models and particle filtering. IEEE Proceedings of Fifth International Conference on Automatic Face and Recognition, 2002, 405-410.

[Careneiro+09] Carneiro, A.; Cortez, P.; Costa, R. Reconhecimento de gestos da libras com classificadores neurais a partir dos momentos invariantes de Hu. Interaction – 09, São Paulo, 2009, 190-195.

[Chen+03] Chen, F.; Fu, Ch.; e Huang, Ch. Hand gesture recognition using a real-time tracking method and hidden Markov models. Image and Vision Computer, 21, (2003), 745-758.

[Chen+07] Chen, Q.; Geoganas, N.D.; Petriu, E.M. Real-time vision-based hand gesture recognition using haar-like features. Instrumentation and Measurement Technology Conference – IMTC, Polonia, (2007).

[Elmezain+09] Elmezain, M.; Al-Hamadi, A.; Appenrodt, J.; Michaelis, B. A hidden Markov model-based isolated and meaningfull hand gesture recognition. World academy of science, engineering and technology, 41, 2009, 393-400.

[Gonçalves+12] Gonçalves, N.; Rodrigues, J.; Costa, S. e Soares, F. Automatic detection of stereotypical motor movements. Procedia Engineering, 47 (2012), pag. 590-593.

[Gonzales+10] Gonzales, R.; Woods, R. Processamento digital de imagens, 3a edição, Pearson, 2010, 624 ps.

[Hu62] Hu, M.K. Visual pattern recognition by moment invariants. IEEE Transactions on Information Theory, v8, n2, 1962, 179-187.

[Junior+07] Junio, N.; Kehl, T; Osorio, F; Jung, C. e Musse, S. R. Animando Humanos Virtuais em Tempo Real usando Visão Computacional e Redes Neurais. SVR 2007, RJ-Brasil, 2007.

[Mahbub+12] Mahbub, U.; Imtiaz, H.; Roy, T. Rahman, S., e Ahad, A. R. A template matching approach of one-shot-learning gesture recognition. Pattern Recognition Letters, 2012.

[Manresa+00] Manresa, C.; Varona, J.; Mas, R.; e Perales, F. Real-time hand tracking and gesture recognition for human-computer interaction. Electronic Letter on Computer Vision and Image Analysis, 0(0):1-7, 2000.

[Mitra+07] Mitra, S. Acharya, T. Gesture recognition: A surveys. IEEE Transactions on Systems, Man, and Cybernetics – Part C: Applications and reviews, v37, n3, 2007.

[Mohandes+12] Mohandes, M.; Deriche, M.; Johar, U.; e Ilyas, S. A signer-independent Arabic sign language recognition system using face detection, geometric features, and a hidden Markov models. Computer and Electrical Engineering, 38, (2012), 422-433.

[Murthy+10] Murthy, G.R.S; e Jadon, R.S. Hand gesture recognition using neural networks. IEEE 2nd International Advance Computing Conference, 2010, 134-138.

[Symeonidis00] Symeonidis, K. Hand recognition using neural networks. Master thesis of Science in multimedia signal processing communications, School of Electronic and Electrical Engineering, Surrey University, 2000.

[Viola+01] Viola, P.; Jones, M. Robust real-time object detectin. Cambridge Research Laboratory Technical Report Series CRL2001/01, 2001, 1-24

[Wu+99] Wu, Y. Huang, T. Vision-based gesture recognition: a review. Gesture-based communication in human-computer interaction, 1999, 103-115.

[Yoon+01] Yoon, H.; Soh, J.; Bae, Y.; e Yang, H. Hand gesture recognition using combined features of location, angle and velocity. Pattern recognition, 34 (2001), 1491-1501.