

Um estudo sobre o uso de um sistema multi-agentes para comunicação baseada no rastreamento ocular

Luisa Helena Junqueira Parreira

Resumo—Este trabalho apresenta um estudo sobre o uso de um sistema multi-agentes, juntamente com técnicas de visão computacional, para comunicação baseada no rastreamento ocular. O foco desse sistema é ser uma ferramenta de comunicação voltada para pessoas com capacidade de interação verbal e gestual restrita ou comprometida, e que podem utilizar apenas seus olhos como forma de interação com o ambiente.

São descritos aqui um breve histórico e conceitos relacionados à área de rastreamento do olhar, alguns conceitos básicos sobre visão computacional e reconhecimento de padrões, definições relacionadas à comunicação visual e aos sistemas multi-agentes que são importantes para o contexto do trabalho. Em seguida, é proposto um modelo de um sistema multi-agentes para comunicação baseada no rastreamento ocular. Por fim, são apresentadas algumas questões a serem analisadas e alguns problemas a serem solucionados considerando o domínio da aplicação tratado nesse trabalho.

Palavras-chave—rastreamento do olhar, sistema multi-agentes, comunicação visual.

I. INTRODUÇÃO

O objetivo deste trabalho é apresentar um estudo sobre o uso de um sistema multi-agentes, juntamente com técnicas de visão computacional, para comunicação baseada no rastreamento do olhar. O foco desse sistema é ser uma ferramenta de comunicação voltada para pessoas com capacidade de interação verbal e gestual restrita ou comprometida, e que podem utilizar apenas seus olhos como forma de interação com o ambiente.

O estudo e utilização de técnicas de rastreamento do olhar no desenvolvimento de sistemas têm crescido nos últimos anos. Tais técnicas permitem a interação do usuário com o sistema sem a exigência do contato físico com o computador, através do mouse e do teclado. Basicamente, através das informações obtidas dos sistemas de rastreamento do olhar, é possível determinar o ponto observado pelo usuário ou estimar sua linha de visão, quando não é considerada a observação de uma superfície em particular.

Essas técnicas são particularmente importantes quando consideramos pessoas com capacidade de comunicação verbal ou gestual restrita ou comprometida, mas que possuem suas

funções cognitivas em perfeito estado. Para essas pessoas, as informações obtidas através das técnicas de rastreamento do olhar podem significar a construção de um canal de comunicação natural e rápido, pois utilizamos nossos olhos para explorar constantemente o ambiente à nossa volta. Nesse caso, um protocolo de comunicação seria definido, possibilitando mapear cada movimento dos olhos para seu respectivo significado em linguagem escrita.

Assim, um sistema composto por agentes de software capazes de captar o movimento dos olhos de um agente humano, decodificar essas informações em linguagem escrita e enviá-las para um dispositivo de saída, poderá beneficiar pessoas que possuem limitações físicas e capacidade de comunicação comprometida, mas que podem ver, ouvir e compreender normalmente o que está acontecendo à sua volta.

Considerando que os olhos são capazes de mudar seu foco de atenção muito rapidamente, o uso de um sistema multi-agentes, nesse contexto, traz a vantagem de possibilitar incorporar essa velocidade ao diálogo com os agentes de software.

Neste trabalho são descritos, inicialmente, um breve histórico e conceitos relacionados à área de rastreamento do olhar. Em seguida, são descritos alguns conceitos básicos sobre visão computacional e reconhecimento de padrões. Na seqüência, são apresentadas algumas definições relacionadas à comunicação visual e aos sistemas multi-agentes que são importantes para o contexto do trabalho. Na última parte, é proposto um modelo de um sistema multi-agentes para comunicação baseada no rastreamento ocular. Por fim, são apresentadas algumas questões a serem analisadas e alguns problemas a serem solucionados considerando o domínio da aplicação tratado nesse trabalho.

II. RASTREAMENTO DO OLHAR

As pesquisas relacionadas ao rastreamento do movimento dos olhos se desenvolveram a partir de 1970, com o avanço das tecnologias para rastreamento do olhar e das teorias psicológicas para relacionar as informações dos movimentos dos olhos aos processos cognitivos. Muitos trabalhos procuravam estudar o funcionamento do olho humano e o que ele poderia revelar sobre percepção e processos cognitivos. A partir de 1980, com a proliferação dos computadores pessoais, desenvolveram-se pesquisas voltadas para a utilização do rastreamento do movimento dos olhos aplicadas ao campo da

interação humano-computador. Desde 1990, com o avanço de tecnologias como, por exemplo, a Internet, e-mail e videoconferência, as pesquisas cresceram no sentido de responder questões sobre usabilidade e sobre a utilização do rastreamento do olhar como um mecanismo de entrada de dados.[4]

Nos últimos anos, o uso da interação humano-computador usando o movimento dos olhos, em tempo real, tem sido, muitas vezes, estudada para aplicações voltadas para usuários com algum tipo de limitação física e que podem utilizar somente seus olhos como *input* de sistemas. Ou seja, as imagens do movimento dos olhos são captadas por um dispositivo de entrada e as informações obtidas a partir dessas imagens são utilizadas pela aplicação em questão.

As técnicas de rastreamento do movimento dos olhos são divididas em duas categorias:

- Rastreamento do olhar intrusivo: técnicas que exigem o contato de algum instrumento de medição com o usuário do sistema. Um exemplo é a técnica conhecida como eletrooculografia. Ela consiste no uso de eletrodos que são colocados ao redor dos olhos e medem a diferença de potencial existente entre a córnea e a retina. Através dessa medição é possível captar movimentos relativos do olho. As técnicas intrusivas causam desconforto e possuem pouca praticidade, devido ao contato com os instrumentos de medição. Esses fatores impedem o uso de rastreamento do olhar intrusivo em um contexto de comunicação, que é o foco desse trabalho.
- Rastreamento do olhar não-intrusivo: técnicas que, através de imagens de vídeo do olho, estimam a posição observada pelo usuário do sistema, baseando-se no reconhecimento de alguma característica do olho. São as técnicas indicadas para o tipo de aplicação considerada neste trabalho.

Alguns exemplos dessa técnica são citados a seguir:

- Identificação da fronteira entre a esclera (parte branca) e a íris (parte colorida). Devido à diferença de tonalidade entre a esclera e a íris, essa fronteira é facilmente identificada. Porém, pelo fato da íris ser parcialmente coberta pela pálpebra, a precisão vertical do rastreamento é bastante limitada.
- Identificação da pupila utilizando uma fonte de luz infravermelho alinhada ao eixo óptico da câmera. A técnica consiste em incidir a luz no olho do usuário do sistema. Assim, a córnea irá refletir a luz, produzindo o efeito conhecido como “olho vermelho”, que produzirá pupilas com alto brilho, em imagens em níveis de cinza. Essa característica (alto brilho) será utilizada para determinar a posição do olho.

- Uma técnica alternativa, também não intrusiva, é o uso de rede neural. Essa técnica não utiliza como base a identificação de alguma característica do olho em particular, como nas técnicas apresentadas acima. Essa técnica utiliza uma rede neural para aprender a relação entre as coordenadas globais e as coordenadas da projeção da imagem na câmera, possibilitando, assim, identificar a posição observada pelo usuário.

O estudo de técnicas de rastreamento de imagem pode ser considerado uma subdivisão da área de visão computacional.

III. VISÃO COMPUTACIONAL E RECONHECIMENTO DE PADRÕES

De forma simplificada, podemos dizer que visão computacional é o conjunto de métodos e técnicas que possibilitam aos sistemas computacionais interpretar imagens. Em outras palavras, podemos dizer que a finalidade da área de visão computacional é obter, a partir de uma imagem digital de entrada, informações úteis e significativas para a aplicação em questão, como por exemplo, reconhecimento de características. Nesse sentido, a utilidade e significado de uma informação obtida através de técnicas de visão computacional é altamente dependente da aplicação.

A organização de um sistema de visão computacional, apesar de não possuir um modelo fixo bem definido para todas as aplicações, possui alguns processos que são comuns em grande parte dos sistemas desenvolvidos nessa área. Não faz parte do objetivo desse trabalho explicar em detalhes os processos que compõem um sistema de visão computacional, mas apenas citá-los resumidamente:

- Aquisição de imagem: consiste em obter uma imagem ou uma seqüência de imagens através de sensores, com a finalidade de adquirir dados de entrada para posterior processamento.
- Pré-processamento: consiste em aplicar métodos de processamento de imagem, que irão auxiliar o processamento de fato, que ocorrerá posteriormente. Como exemplo podemos citar: rotacionar a imagem, redução da resolução espacial, entre outros.
- Segmentação de imagem: é um processo utilizado para retirar da imagem processada aquilo que não é de interesse da aplicação, mantendo-se apenas o objeto de interesse. Consiste em fragmentar uma determinada imagem em regiões ou objetos, baseando-se em suas características (cor, intensidade, etc) ou no formato dos objetos (contornos, pontos, descontinuidades, etc). Assim, uma determinada região da imagem será destacada, segmentada e essa informação será armazenada para futuro processamento.
- Extração de características: consiste em identificar informações de uma imagem, ou seja, a partir da imagem já processada pela etapa de segmentação, são extraídos desses objetos características que

servam para identificá-los, como: tamanho, textura, bordas, cantos, etc.

- Classificação: consiste na classificação dos objetos detectados e processados através das etapas descritas acima, em diferentes categorias ou classes pré-definidas.

As etapas de segmentação, extração de características e classificação, são normalmente encontradas em sistemas de visão computacional envolvendo reconhecimento de padrões.

Reconhecimento de padrões é uma etapa importante em grande parte das aplicações de visão computacional, e tem como objetivo o uso de teorias e técnicas que permitem a classificação de objetos, considerando um conjunto de categorias ou classes. Esses objetos não são apenas imagens, mas podem ser também, sons, seqüência de caracteres ou qualquer outro tipo de sinal (digitalizado), que possa ser capturado através de sensores.

Um sistema para reconhecimento de padrões é composto, basicamente, por um sensor que obtém as imagens ou outros objetos a serem processados, um mecanismo de extração de características, que processa as informações numéricas ou simbólicas extraídas dos objetos, e um esquema de classificação dessas informações, que dependerá das características que forem identificadas nos respectivos objetos.

O esquema de classificação dos objetos em categorias ou classes, ou seja, o reconhecimento de padrões, pode ocorrer basicamente de duas formas:

- Baseado em um conjunto de padrões que são disponibilizados *a priori*.
- O sistema não recebe informações anteriores sobre os padrões a serem considerados, de forma que estabelece as classes ou categorias através de análise de padrões probabilísticos.

No caso da aplicação tratada nesse trabalho, o reconhecimento de padrões, a partir do rastreamento do olhar, é fundamental. O sistema irá identificar nas imagens captadas e processadas, os padrões de movimento dos olhos ou de fixação em algum ponto e mapear essas informações para as respectivas classes de movimento (coordenadas de observação). Esses padrões identificados serão decodificados em linguagem escrita pelos agentes do sistema e enviada para um dispositivo de saída, conforme será explicado nas próximas seções.

IV. COMUNICAÇÃO VISUAL

Comunicação visual é todo canal de comunicação expresso com a utilização de componentes visuais, como: imagens, desenhos, símbolos, gráficos, ou seja, tudo o que pode ser visto.

Neste trabalho, o termo comunicação visual é utilizado para definir o canal de comunicação que utiliza os olhos e as pálpebras para expressar pensamentos, idéias, necessidades, ou seja, como canal de comunicação.

Para que haja a comunicação visual, é necessário estabelecer um protocolo de comunicação, que irá definir as

regras para a decodificação das mensagens enviadas através do movimento dos olhos do agente humano.

As figuras a seguir mostram dois exemplos de placas utilizadas para a comunicação visual de pessoas com capacidade de comunicação verbal e gestual restrita ou comprometida.

	1	2	3	4	5	6
1	A	B	C	D	E	F
2	G	H	I	J	K	L
3	M	N	O	P	Q	R
4	S	T	U	V	W	X
5	Y	Z	1	2	3	4
6	5	6	7	8	9	0
7	Ç	SEDE	FOME	FRIO	CALOR	LUZ

Fig. 1 Cartela básica para comunicação visual

A Figura 1 mostra uma cartela básica para comunicação visual. Nela estão numeradas as linhas e as colunas. O cruzamento entre cada linha e coluna identifica uma letra ou um número. A pessoa deve sinalizar a linha e a coluna selecionadas, cujo cruzamento contém a informação que deseja. Isso pode ser feito, piscando os olhos, por exemplo, ou de uma outra forma, previamente combinada. Assim, estará sinalizando as letras ou números que irão compor cada palavra ou frase que deseja comunicar. As informações na última linha da cartela devem ser adaptadas às necessidades individuais de cada pessoa.

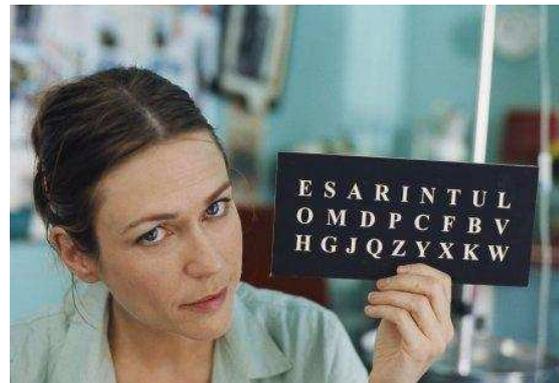


Fig. 2 Placa para comunicação visual do filme "O Escafandro e a Borboleta".

A Figura 2 mostra um outro modelo de placa para comunicação visual, cujas letras que irão compor a mensagem também devem ser selecionadas com o piscar dos olhos ou alguma outra forma definida anteriormente.

Independente do modelo de comunicação visual utilizado é essencial a criação de um protocolo de comunicação, que possibilitará a definição de regras para a decodificação das mensagens. Por exemplo, quando uma mensagem é iniciada ou

finalizada, o que significa piscar os olhos, entre outras definições.

V. SISTEMA MULTI-AGENTES

No estudo apresentado nesse trabalho, é considerado um sistema formado por um conjunto de agentes capazes de cooperar entre si na realização de suas tarefas. É um sistema composto por um agente humano que envia mensagens, através do movimento dos olhos, para serem processadas pelos agentes de software.

É pressuposto que os agentes desse sistema, como qualquer agente, devem ter a capacidade de perceber o ambiente em que estão situados e agir sobre ele, em resposta a mudanças ocorridas nesse mesmo ambiente. No contexto de um sistema para comunicação baseada no rastreamento do movimento dos olhos, essa propriedade de reatividade significa a atenção permanente dos agentes de software à movimentação dos olhos do agente humano, que poderá gerar uma mensagem a ser decodificada e enviada a um dispositivo de saída.

Assim, a comunicação entre os agentes estará ocorrendo através da interação entre eles, ou seja, através da troca de mensagens que resultará em uma ação, como resposta à mensagem enviada. Conceitualmente, interação entre agentes não é sinônimo de comunicação entre agentes, já que para haver comunicação de fato é necessário a implementação de uma Linguagem de Comunicação de Agentes (ACL). Não faz parte do objetivo desse estudo, descrever maiores detalhes sobre ACL, já que os agentes do sistema em questão não são agentes comunicativos, havendo somente interações entre eles, resultando em colaborações.

Entretanto, é utilizado nesse estudo o termo comunicação de forma genérica, ou seja, considerando que um agente ao emitir uma mensagem a outro agente, está realizando uma forma de comunicação, na medida em que a mensagem ao ser interpretada pelo agente receptor, irá resultar em uma ação ou comportamento como resposta.

Entre tantas definições de agentes que existem, destacamos a seguir duas que se enquadram ao contexto da aplicação em questão:

- *“O termo agente é utilizado para representar dois conceitos ortogonais. O primeiro é a habilidade de execução autônoma e o segundo é a habilidade em domínios específicos” (Sankar Virdhagriswaran – Crystaliz Inc. – MuBot).*

Essa primeira definição descreve duas propriedades particularmente importantes para os agentes em um sistema de comunicação baseada no rastreamento do olhar.

Uma propriedade é a autonomia, que possibilita ao agente ter o controle sobre suas próprias ações, sem precisar de autorização para executá-las. Aplicada ao contexto do trabalho, essa propriedade define que o agente humano poderá, a qualquer momento, iniciar a comunicação visual. Isso significará aos agentes de software as tarefas de rastreamento do movimento dos olhos do agente humano, posterior decodificação da mensagem em linguagem escrita e seu envio a um dispositivo de saída. Todo esse processo, desde o início

da comunicação pelo agente humano até o envio da mensagem decodificada ao dispositivo de saída, é realizado de maneira autônoma e independente.

A outra propriedade descrita nessa primeira definição é a habilidade em domínios específicos. Na introdução, foi explicitado que o sistema, objeto de estudo desse trabalho, tem como foco ser uma ferramenta de comunicação voltada para pessoas com capacidade de interação verbal e gestual restrita ou comprometida. Pessoas nessa situação costumam comunicar-se através da visão seguindo um protocolo de comunicação, conforme explicado na seção anterior. Portanto, a propriedade de possuir habilidade em domínio específico irá significar que, todos os agentes que atuam nesse sistema, conhecem o protocolo de comunicação estabelecido para o contexto específico da aplicação.

- *“Agentes de software são programas que se empenham em diálogos de forma a negociar e coordenar a transferência de informação” (Michael Coen – MIT AI Lab – SodaBot).*

Para que o sistema multi-agentes realize, efetivamente, a conversão da comunicação visual para a comunicação escrita, é necessária a troca freqüente de mensagens entre os agentes para possibilitar a cooperação entre eles e, assim, tornar possível a transferência das informações para processamento de forma coordenada. Nesse caso, o agente humano estará, através do movimento dos olhos, enviando mensagens aos agentes de software do sistema. Essas mensagens irão significar um conteúdo que o agente humano deseja comunicar. Ou seja, o agente humano espera que as mensagens por ele enviadas sejam decodificadas pelos agentes de software, para posterior comunicação em linguagem escrita.

VI. UM SISTEMA MULTI-AGENTES PARA COMUNICAÇÃO BASEADA NO RASTREAMENTO OCULAR

O uso de um sistema multi-agentes em uma aplicação de rastreamento do movimento dos olhos voltada para comunicação traz algumas vantagens, como:

- Considerando que o sistema estará atuando em um ambiente dinâmico, possibilita a atenção permanente dos agentes, podendo identificar o início da comunicação a qualquer instante.
- Agiliza o processo de comunicação, na medida em que cada agente é responsável por uma parte do processo, podendo atuar simultaneamente.
- Considerando que os olhos são capazes de mudar seu foco de atenção muito rapidamente, possibilita ao agente humano incorporar essa velocidade ao diálogo com os agentes de software.

O sistema multi-agentes, proposto nesse trabalho, é composto por um agente humano e agentes de software que interagem e cooperam entre si.

É um sistema formado por agentes especializados, onde cada agente é especializado em uma determinada tarefa. Ou

seja, cada agente possui potencialidades específicas. Assim, o agente humano é o responsável por uma única tarefa, que é enviar a mensagem visual (informação) aos demais agentes, e os agentes de software são responsáveis por processar essa informação. Cada etapa do processamento da informação é executada por um agente de software específico.

O modelo de operação do sistema multi-agentes proposto é a coordenação centralizada, que significa que um dos agentes é o “agente central”, que envia os comandos para os outros agentes executarem. Nesse caso, o agente central é o agente humano, pois é ele que envia as mensagens que resultam em ações dos demais agentes do sistema. Os agentes de software ficam aguardando o agente humano iniciar a comunicação e, a partir dela, iniciam a execução de suas tarefas.

Os agentes de software que compõem o sistema são agentes reflexivos. Nesse tipo de agente, a ação que ele executa é uma função direta da sua percepção em relação ao ambiente. Ou seja, as ações dos agentes de software estão diretamente relacionadas com o tipo de informação coletada pelo rastreamento do movimento dos olhos do agente humano.

É apresentado na Figura 3, o modelo do sistema multi-agentes proposto nesse trabalho. É importante lembrar que cada agente proposto nesse modelo pode ser decomposto em sub-agentes, se for necessário para a execução das respectivas tarefas.

A seguir, são descritas as funções de cada agente dentro do sistema, conforme mostrados na Figura 3:

Função do Agente humano

- Executar a comunicação visual: conforme explicado na seção IV, é a forma de comunicação que utiliza os olhos como meio para realizar essa comunicação. Assim, através do movimento dos olhos e, conforme um protocolo de comunicação estabelecido, o agente humano irá enviar as mensagens para os agentes de software.

Função do Agente A: Rastreador do olhar

Utilizando alguma técnica de rastreamento não-intrusivo, como explicado na seção II, o agente A irá:

- Adquirir as imagens e processá-las: estará constantemente recebendo as imagens captadas pelo sensor. As imagens serão processadas, conforme explicado na seção III.
- Reconhecer os padrões: Mapear as imagens processadas para as respectivas classes de movimentos. Conforme explicado na seção IV, o movimento dos olhos, a fixação dos olhos em um ponto ou o piscar dos olhos, podem significar uma determinada informação, de acordo com o protocolo de comunicação estabelecido. Os padrões identificados serão enviados para um buffer.

Função do Agente B: Coletor

- Identificar a presença de mensagens a serem decodificadas: estará constantemente verificando no

buffer a existência de mensagens processadas pelo Agente A. Conforme o protocolo de comunicação definido, o Agente B será capaz de identificar o início e o fim de uma sequência de mensagens, ou seja, o início e o fim de uma frase e efetuar sua transferência para o Agente decodificador. As mensagens serão transferidas para um buffer, onde ficarão até serem processadas pelo Agente C (decodificador).

Função do Agente C: Decodificador

- Decodificar as mensagens: estará constantemente verificando no buffer a existência de mensagens processadas pelo Agente B. Considerando os padrões identificados pelo Agente A, irá decodificar cada mensagem encontrada no buffer para seu correspondente caractere do alfabeto português.
- Enviar a mensagem decodificada para o dispositivo de saída especificado: a mensagem, já decodificada em linguagem escrita, será enviada para um dispositivo de saída, que poderá ser um monitor, um endereço de e-mail, entre outros.

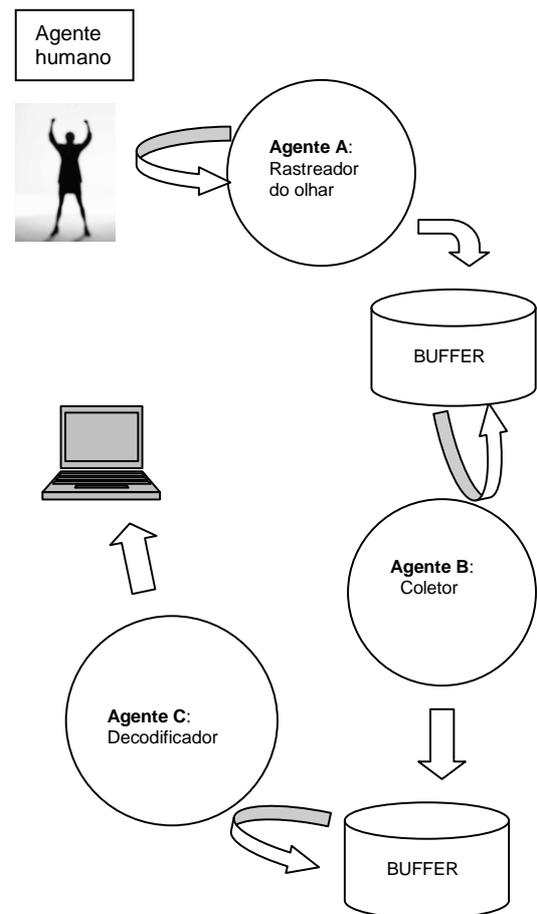


Fig. 3 Modelo

VII. CONSIDERAÇÕES FINAIS

Este trabalho apresentou um breve estudo sobre o uso de um sistema multi-agentes, juntamente com técnicas de visão computacional, para comunicação baseada no rastreamento do olhar. O sistema estudado teve como foco principal ser uma ferramenta de comunicação voltada para pessoas com capacidade de interação verbal e gestual restrita ou comprometida, e que podem utilizar apenas seus olhos como forma de interação com o ambiente. Esse sistema pode ser particularmente importante para essas pessoas, no sentido de possibilitar um canal de comunicação natural e rápido com o ambiente. Isso poderá significar, em muitos casos, uma possibilidade de inserir-se novamente ao convívio social, na medida em que torna possível a interação com outras pessoas.

É importante destacar algumas questões a serem aprofundadas ao considerar a implementação de um sistema como esse.

Uma primeira questão é que nem sempre é possível saber, com exatidão, o ponto focado pelo olho. Isso porque as técnicas não-intrusivas, que são indicadas para a aplicação estudada nesse trabalho, muitas vezes, possuem precisão limitada.

Outra questão a ser considerada em uma implementação, é que, muitas vezes, uma seleção é ativada em qualquer posição observada, sem a intenção do usuário. Uma solução possível é considerar a seleção de um ponto específico somente após um intervalo de tempo. Porém não é a única solução e precisa ser devidamente analisada.

Uma terceira questão importante é que as técnicas de rastreamento não-intrusivas necessitam de um procedimento de calibragem. Isto é, mapear a posição de determinada característica na imagem de vídeo captada em relação à posição da linha de visão do indivíduo. Cada vez que o agente humano é movimentado esse procedimento precisa ser realizado novamente. Existem várias técnicas de calibragem que precisam ser analisadas, considerando a usabilidade do sistema. Existem estudos relacionados às técnicas de rastreamento do olhar que não exigem calibragem e que obtêm razoável precisão. Na implementação do modelo será necessário estudar com maior profundidade essa questão.

REFERÊNCIAS

- [1] A. T. Duchowski, “A breadth-first survey of eye-tracking applications”, 2002.
- [2] A. T. Duchowski, “Eye tracking methodology: theory and practice”, 2006.
- [3] R. R. Pinho, J.M. Tavares, M. Correia, “Introdução à análise de movimento usando visão computacional”, 2004.
- [4] R. J. Jacob, K. S. Karn, “Eye Tracking in Human-Computer interaction and Usability Research: Ready to Deliver the Promises”, 2002.
- [5] K.P. Souza, H. Pistori, “Implementação de um Estrator de Características baseado em momentos da imagem”, 2004.
- [6] T. F. Vieira, E. Fontana, “Dispositivo de rastreamento de movimentos oculares baseado em Webcam e iluminação com led infravermelho”, 2006.

- [7] K. Kim, R. S. Ramakrishna, “Vision-based eye-gaze tracking for human computer interface”, 1998.
- [8] M. Wooldridge, “An introduction to multiagent systems”, 2002.
- [9] A. P. Pinheiro, “Desenvolvimento de um sistema de captura e análise de movimentos baseado em técnicas de visão computacional”, 2008.
- [10] J. A. França, M. R. Stemmer, “Stereo calibration with a free-moving stick”, 2006.
- [11] R. S. Guerra, “Calibração automática de sistemas de visão estéreo a partir de movimentos desconhecidos”, 2004.