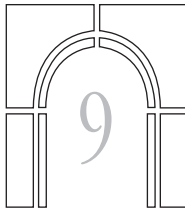

Cómo citar este artículo: Uribe, A. y Alves dos Reis, S. (2015, enero-junio). Rastreamento de rosto como ferramenta interativa e de monitoramento do estado emocional do usuário. Rev. Cient. Gen. José María Córdova 13(15), 245-255



Rastreamento de rosto como ferramenta interativa e de monitoramento do estado emocional do usuário*

Recibido: 7 de enero de 2015 • Aceptado: 1 de febrero de 2015

Rastreo de facciones como herramienta interactiva y detección del estado de ánimo del usuario

System to Track Factions as an Interactive and Monitoring Tool for the User's Particular Style

Système de suivi des factions comme un outil interactif en ligne et de surveillance de l'état d'esprit de l'opérateur

Alvaro Uribe- Quevedo ^a

Silas Alves dos Reis ^b

* Este artigo é resultado de uma colaboração investigativa entre a Universidade Militar Nueva Granada e la Universidade de São Paulo (Este artículo es resultado de una colaboración investigativa entre la Universidad Militar Nueva Granada y la Universidade de São Paulo).

^a Universidad Militar Nueva Granada, Bogotá, Colombia. Doutor em Engenharia Mecânica. Contacto: alvaro.uribe@unimilitar.edu.co

^b Universidade de São Paulo, São Carlos, Brasil. Mestre em Engenharia Mecânica. Contacto: silasalves@gmail.com

Resumo. Um dos principais meios de comunicação é o rosto, através do qual, as pessoas são capazes de comunicar – voluntaria ou involuntariamente – informação a respeito de seu estado emocional, tal como alegria, tristeza, raiva ou medo. A face humana já era objeto de estudo antes do advento da área de visão computacional, entretanto o recente crescimento da visão de máquina resgatou o interesse no estudo da face para detectar o estado emocional do usuário. Atualmente, existem várias alternativas para a detecção de expressões do rosto baseadas em imagens bidimensionais ou matrizes de profundidade. Ademais, a massificação de interfaces de usuário tridimensionais (3DUIs), impulsionada principalmente pela indústria de videogames, viabilizou o uso de gestos corporais como fonte de informações do usuário. Neste projeto, desenvolveu-se um aplicativo para detectar expressões faciais que indiquem desconforto dentro de um cenário de trabalho. Seu objetivo é avaliar a detecção de expressões por intermédio de sensores baseados em visão de máquina e cuja funcionalidade está baseada no reconhecimento de formas tridimensionais. Uma análise comparativa foi realizada entre dois tipos de sensores de fácil aquisição no mercado com a finalidade de avaliar seu potencial em um cenário do cuidado da saúde ocupacional.

Palabras-chave: expressões faciais como fonte de informações do usuário, ferramenta interativa, interfaces de usuário tridimensionais, monitoramento do estado emocional do usuário, rastreamento de rosto.

Resumen. El rostro humano es uno de los canales de comunicación que permiten identificar de forma voluntaria o involuntaria información de los estados emocionales del usuario, como lo son el miedo, la rabia, la felicidad o la alegría. Las expresiones faciales han sido objeto de estudio desde antes de que la visión por computador fuese un área de interés; sin embargo, los avances recientes de la visión artificial han renovado el interés por el seguimiento digital de facciones del rostro humano para detectar los estados emocionales del usuario. En la actualidad las técnicas más comunes de seguimiento facial utilizan imágenes 2D y mapas de profundidad, pero debido a la masificación de interfaces de usuario 3D y gracias al auge de los videojuegos, es posible contar con sensores que realizan captura de movimiento con diferentes finalidades y prestaciones. En este proyecto desarrollamos una aplicación para la detección de expresiones faciales relacionadas con la incomodidad, en un escenario de trabajo frente al computador. Su objetivo es validar las expresiones del rostro, utilizando sensores de detección de visión artificial a partir de la forma tridimensional del rostro. Para validar los resultados realizamos un análisis comparativo utilizando sensores comerciales para valorar su potencial en el contexto de conservación y prevención de la salud ocupacional.

Palabras claves: detección del estado emocional del usuario, expresiones faciales como una fuente de información del usuario, herramienta interactiva, interfaces de usuario 3D, rastreo de facciones.

Abstract. The human face is one of the main communication mean that allows conveying– voluntary or involuntary – information about emotions such as, happiness, sadness, anger or fear. The face expressions have been studied before computer vision became a wide known area of research; however, the evolution of machine vision restarted the interest on face tracking to detect user emotions. Current face tracking techniques use 2D images and depthmaps that in conjunctions with recent advances in 3D user interfaces mainstreamed due to videogames are providing motion tracking tools to gather interaction information from the user. In this project we developed a software application that detects facial expressions related do discomfort within a computer-based working scenario. The goal of this development is to validate the tracking of such expressions using machine vision based on the 3D shape of the face. To validate the results, we performed a comparative analysis between affordable sensors in the market to assess their potential within an occupational health care scenario.

Keywords: face Tracking, facial expressions as a source of user information, interactive tool, monitoring the emotional state of the user, three-dimensional user interfaces

Résumé. Le visage humain est l'un des canaux de communication qui permettent d'identifier l'information des états volontairement ou involontairement utilisateur émotionnels, comme la peur, la colère, le bonheur ou la joie. Les expressions faciales ont été étudiées depuis avant la vision par ordinateur étaient un sujet de préoccupation, cependant, les progrès récents dans la vision de la machine ont un regain d'intérêt dans les fonctions numériques de suivi du visage humain de détecter les états émotionnels utilisateur. Aujourd'hui, les techniques les plus communs utilisés suivi du visage images en 2D et des cartes de profondeur, mais en raison de la masse des interfaces utilisateur 3D grâce à la hausse de jeux vidéo, vous pouvez avoir des capteurs qui effectuent la capture de mouvement à des fins différentes et des avantages. Dans ce projet, nous avons développé une application pour la détection des expressions faciales liées à l'inconfort, de travailler sur une scène devant des ordinateurs. Son objectif est de valider les expressions du visage, en utilisant des capteurs pour détecter la vision artificielle de la forme tridimensionnelle du visage. Pour valider les résultats effectué une analyse comparative en utilisant des capteurs commerciaux afin d'évaluer leur potentiel dans le contexte de la conservation et de la prévention de la santé au travail.

Mots-clés: expressions faciales comme une source d'information de l'utilisateur, interfaces utilisateur 3D outil interactif, suivi de factions, surveillance de l'état émotionnel de l'utilisateur.

Introdução

O computador é, sem sombra de dúvidas, o maior ícone do mundo moderno. Ele está presente na sociedade nas mais diversas formas – como computadores de mesa, *tablets*, telefones celulares (*smartphones*), e até mesmo em geladeiras e relógios (*smartwatches*). Sua versatilidade aperfeiçoou e acelerou processos industriais, revolucionou os meios de comunicação e, principalmente, transformou o modo com o qual as pessoas trabalham e colaboram entre si. Estas vantagens propiciadas pela sua adoção, no entanto, não vêm desacompanhadas de custos.

Uma causa comum das enfermidades ocupacionais está relacionada a problemas musculoesqueléticos que reduzem a funcionalidade das pessoas. No caso do trabalho relativo ao uso do computador, ou outros sistemas computacionais, o sedentarismo apresenta um fator de alto risco, pois o trabalhador é obrigado a manter a mesma postura durante longos períodos, podendo resultar em esforços excessivos das estruturas musculares devido à rigidez ou em atrofiamento devido à falta de uso certos grupos de músculos (WHO, 2014).

Neste contexto, sistemas que sejam capazes de medir o tempo de exposição do usuário de computadores a más posturas, ou que indique seu estado emocional, podem fornecer informações essenciais sobre a saúde do trabalhador. Várias abordagens podem ser adotadas para a criação destes sistemas e, dentre elas, encontra-se o rastreamento do rosto e reconhecimento de emoções faciais.

O rosto humano tem sido objeto de estudos com o intuito de compreender aspectos fisiológicos e psicológicos do comportamento das pessoas durante a interação com seu entorno baseando-se nas expressões faciais realizadas (Paul, 2005). Como resultado do entendimento do funcionamento do sistema visual humano, a visão de máquina tornou-se objeto de pesquisa com a finalidade de criar sistemas computacionais capazes de adquirir imagens de câmeras e processá-las para obter informação de uma cena sem a intervenção humana (Schalkoff, 1989). Com avanço das tecnologias de eletrônica e mecânica, a miniaturização de componentes e o aumento da po-

tência de processamento tem permitido o desenvolvimento de sistemas cada vez mais completos para aplicações diversas, tais como a inspeção de circuitos (Sanz & Jain, 1986), monitoramento em células robóticas (Ye & Tsotsos, 1999), revisão de dispositivos médicos (Lahajnar, Pernuš, & Kovačič, 2002), e detecção de rosto humano (Yow & Cipolla, 1997), dentre outros.

Na área de detecção e rastreamento do rosto humano, os interesses de pesquisa têm evoluído conforme a qualidade dos dispositivos de captura de imagem é aperfeiçoada e a capacidade de processamento dos dispositivos de computação tem aumentado, alcançando inclusive e aplicações para dispositivos móveis, tais como *smarthphones* e *tablets*, em cenários como o aumento da visão periférica (Yang, Hasan, Bruce, & Irani, 2013). Outra área de aplicação utilizando rostos é a de sistemas biométricos, empregados na segurança de acesso em prédios ou computadores (Samal & Iyengar, 1992).

Tão logo as tecnologias de imagem digital reduziram seu custo e se popularizaram, como a webcam, os trabalhos sobre rastreamento de rosto começaram a aumentar e, assim, foram identificadas outras áreas de aplicação, como interfaces de usuário para pessoas com deficiência motora para manipulação de robôs móveis (Tsai & Song, 2006), detecção de desordens do sono e problemas de saúde ocupacional (Perez, Palma, Holzmann, & Pena, 2001), interfase para videojogos (Wang, 2006), e inclusive para reabilitação baseadas em gestos do rosto (González-Ortega, Antón-Rodríguez, Díez-Higuera, & Boto-Giralda, 2010).

O interesse em interfaces para pessoas com algum tipo de deficiência é resultado do crescimento do número destas pessoas, que em dezembro de 2014 chegou a cerca de 15% da população mundial, com um número entre 110 milhões e 190 milhões de adultos com dificuldades para levar uma vida completamente funcional (WHO, Disability and health, 2014), devido a isto, a Organização Mundial da Saúde tem criado iniciativas que procuram subsidiar melhores condições para que as pessoas com deficiência disfrutem de boa qualidade de vida. Uma iniciativa é a saúde eletrônica (*e-health*), a qual consiste em disponibilizar soluções baseadas em tecnologia para melhorar a vida das pessoas com deficiência (WHO, eHealth, 2014). Neste cenário, o desenvolvimento de interfaces de usuário tridimensionais (3DUI) tem oferecido uma alternativa de dispositivos que podem ser utilizados para múltiplos propósitos, tanto ao entretenimento quanto ao cuidado da saúde, pois capturam dados de movimentação no espaço 3D (Wingrave, et al., 2010) (Francese, Passero, & Tortora, 2012).

Tradicionalmente, os algoritmos de rastreamento de face e detecção de expressões baseiam-se em sequências de imagens 2D, porém a necessidade de dados que descrevessem as geometrias da face no espaço 3D culminou no uso de novas abordagens, como o emprego da visão estereoscópica para reconstituir a face tridimensionalmente (Matsumoto & Zelinsky, 2000). No entanto, os dispositivos capazes de reconstituir o espaço 3D só foram massificados e popularizados na última década pelos consoles para videojogos, facilitando o acesso a sensores como o Kinect® da Microsoft que, segundo a fabricante, vendeu aproximadamente 24 milhões de unidades desde seu lançamento (Mehdi, 2013).

Desta forma, considerando a quantidade de tempo que os usuários estão frente a sistemas computacionais e o risco de desenvolvimento de problemas de saúde, este trabalho propõe analisar a aplicação de uma 3DUI como ferramenta para identificar o estado emocional do usuário por meio das expressões faciais, de modo que ela permita avaliar se uma interação gera algum

desconforto que possa resultar num problema físico. Neste sentido, o sensor escolhido será o Microsoft Kinect. O desenvolvimento será realizado considerando as características do sensor relativas à captura de informação 3D do usuário (Zhang, 2012), realizando-se comparativos entre a primeira e a segunda versão do sensor, para gerar alertas que permitam a um especialista da saúde tomar medidas preventivas ou corretivas de acordo com o caso.

O presente artigo está organizado em seções, de forma que a Seção 2 volta-se ao desenvolvimento do aplicativo a partir da caracterização do rosto humano e as expressões comuns para indicar estados de conforto e desconforto; a Seção 3 apresenta os resultados obtidos da implementação e análise comparativo dos dados reconhecidos pelos sensores; e finalmente a Seção 4 expõe as conclusões e discussão sobre os resultados são apresentados.

Desenvolvimento

Para o desenvolvimento do aplicativo, é necessário primeiro caracterizar o rosto humano com as expressões que facilitam a identificação de estados que causem desconforto. Estas expressões permitem definir as entradas do sistema, projetar os subsistemas e realimentar estas informações ao usuário através de saídas audiovisuais.

Caracterização do rosto humano

O rosto humano tem a capacidade de expressar numerosas emoções, como a surpresa, raiva, tristeza e felicidade, que são resultado da ação dos músculos da face, nas sobrancelhas, olhos, boca e nariz entre outros. Um resultado evidente são os gestos e os microgestos em combinações que podem ser difíceis de avaliar por sua complexidade (Ekman & Friesen, 2005). Os gestos têm grande importância como meio de comunicação e são objeto de estudo na psicoterapia, onde o paciente exprime seu estado através deles, ou na fisioterapia, que os utiliza para determinar se o cuidado ou tratamento está afetando o paciente. Também são usados por advogados para detectar mentiras ou ocultação de informação por uma testemunha, e até por professores que mudam a estratégia pedagógica segundo a percepção que têm dos estudantes. A informação reconhecida do rosto pode ser utilizada na avaliação de postos de trabalho confortáveis, apresentações em público, videojogos, experiência de usuário, entre outros (Microsoft, 2013).

As expressões do rosto podem ser controladas ou incontroladas. Dentro desse conjunto, existem também expressões voluntárias e involuntárias. Há situações que requerem o uso de expressões controladas como, por exemplo, numa palestra quando se quer reforçar a transmissão de uma mensagem. Por outro lado, há também ocasiões onde se sobressaem expressões involuntárias como o medo, a tristeza, a surpresa ou a raiva. A identificação desses gestos requer a análise do movimento muscular da face. Por exemplo, no caso da expressão da raiva devem se apresentar mudanças em três regiões do rosto: as sobrancelhas descem e se aproximam, as pálpebras e os lábios se tencionam (ou, de acordo com o, caso a boca se abre ou fecha de acordo o nível de raiva, que pode resultar em gritos ou ataque físico) (Ekman & Friesen, 2005).

Arquitetura do sistema

Tomando como base a informação do rosto em termos das expressões que indiquem desconforto, este trabalho considera que as expressões de medo, aversão e raiva permitem conhecer o estado emocional exprimido pelo usuário consciente ou inconscientemente frente a uma atividade realizada no computador. A detecção das expressões é realizada a partir das propriedades disponíveis no *Kit* de Desenvolvimento de Software (SDK) do Kinect, o qual gera uma malha 3D representando o rosto do usuário a partir do mapa de profundidade. A malha permite ao sistema conhecer em que direção os músculos se deslocaram, possibilitando a determinar qual gesto o usuário exprimiu frente a uma Base de Dados (BD).

Para a captura dos gestos serão utilizadas duas 3DUI —Microsoft Kinect e Kinect 2.0— que permitam identificar a variação de movimento no rosto do usuário. Os dados capturados alimentarão o subsistema de processamento de dados, que efetuará seu registro, comparação e armazenamento com o intuito de gerar relatórios que indiquem a expressão detectada em cada momento. A arquitetura do sistema proposto é apresentada na Figura 1.

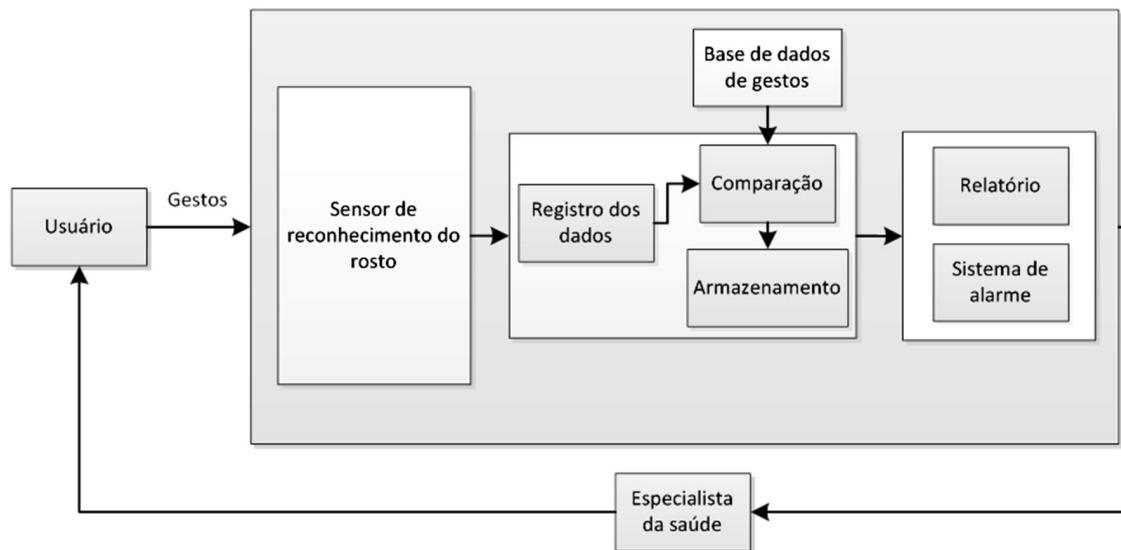


Figura 1. Arquitetura do sistema

Fonte: Elaboração própria.

Características do Sensor

O Kinect necessita do Sistema Operacional (SO) Microsoft Windows 7 ou superior e do Kinect SDK 1.8, este último sendo o fornecedor das classes em linguagem C++ e C# referentes ao rastreamento de face e de geração das malhas 3D do rosto, utilizadas na implementação do sistema. Contudo, devido ao uso do Kinect 2.0, também são necessários o SO Windows 8.1 ou superior. Aplicativos separados foram desenvolvidos para o Windows 7 e 8.1, devido a dados de uso no nível mundial; segundo *NetMarketShare* (Share, 2015) cerca de 55.92% dos computadores de mesa fazem uso do Windows 7, seguido pelo Windows XP com 18.93% e Windows 8.1 com 10.04%.

O desenvolvimento utilizando Kinect toma os dados da malha 3D gerada pelo SDK a partir de 100 pontos reconhecidos de estruturas significativas como as sobrancelhas, olhos, boca e nariz (). Para a detecção das expressões o, SDK 1.8 (Kinect) fornece seis unidades de animação (AUs) para identificar uma face neutra pelos movimentos das sobrancelhas, mandíbula e lábios, ao passo que o SDK 2.0 fornece 17 AUs para o mesmo fim. Além das AU, o SDK fornece unidades de forma (SUs), onze na versão 1.8 e noventa e quatro na 2.0, que descrevem a malha de acordo com as características faciais do usuário. Considerando o caso deste projeto, cujo foco está nas expressões que indicam desconforto, a Tabela 1 apresenta a configuração de unidades.

Tabela 1. Unidades de animação para as expressões de interesse

AU	Expressão
AU5=-1, AU4=+1	Tristeza
AU0=1, AU3=+1	Raiva
AU2=-0.5, AU3=+1	Aversão
AU5=-1, AU2=-0.5, AU0=1	Medo

Fonte: Elaboração própria.

Os pontos de interesse para gerar as alertas são apresentados na Figura 2, em conjunto com as detecções das expressões de interesse.

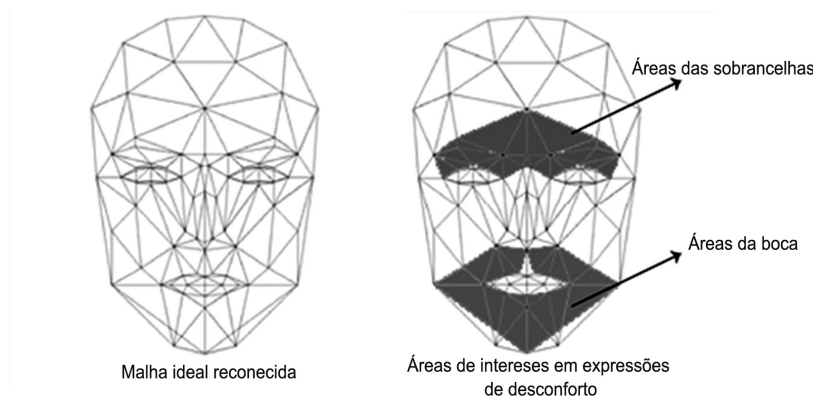


Figura 2. Segmento da malha reconhecida e áreas de interesse

Fonte: Elaboração própria com base em informações.

A chegada da nova versão do sensor, o Kinect 2.0, trouxe uma série de melhorias de hardware – além do aperfeiçoamento do SDK. Originalmente, a primeira versão do Kinect emitia um feixe de luz estruturada no ambiente e reconstituía o mapa de profundidade ao reconhecer as deformações do padrão projetado no ambiente, ao passo que o Kinect 2.0 utiliza um sensor de tempo de voo, resultando em maior resolução e precisão. Para elucidar as diferenças entre as duas versões, a Tabela 2 apresenta um comparativo entre elas.

Tabela 2. Comparativo de características entre o Kinect e o Kinect 2.0

Característica	Kinect	Kinect 2.0
Resolução da câmera colorida	640x480, 30fps	1920x1080, 30fps
Resolução da câmera profundidade	320x240	512x424
Máxima distância de profundidade	~4,5 m	~4,5 m
Mínima distância de profundidade	40 cm modo perto	50 cm
Campo de visão horizontal	57°	70°
Campo de visão vertical	43°	60°
Número de articulações	20	26
Número de esqueletos reconhecidos	2	6
Sistema Operacional	Windows 7 e 8	Windows 8

Fonte: Elaboração própria com base em (Amon & Fuhrmann, 2014).

As capacidades de hardware da versão 2.0, junto ao seu SDK, resultam em funcionalidades adicionais, como a detecção do estado dos olhos (aberto ou fechado), detecção de presença de óculos, rastreamento do olhar e, também, o casamento mais preciso dos pontos de interesse da malha 3D às geometrias do rosto do usuário devido à maior resolução das câmeras.

O aplicativo tem por objetivo fazer rastreamento do rosto e a detecção das expressões e gerar mensagens de alerta apenas nos momentos em que é detectada uma das três expressões de desconforto. Com estas mensagens, o usuário estará ciente da atividade que realizava e poderá relatar a um especialista da saúde os valores a relação entre o evento e a expressão capturada.

Resultados

Obteve-se um aplicativo de rastreamento de face para detectar três cenários de desconforto de acordo com os movimentos musculares associados às sobrancelhas e área da boca. O sistema foi testado com 10 usuários cujo trabalho requer de mais de 6 horas de trabalho contínuo no computador. Durante o teste, não se apresentaram expressões de desconforto devido às atividades realizadas, portanto os participantes realizaram de forma voluntária as expressões que poderiam indicar esse estado.

Nos testes envolvendo o Kinect, a detecção de expressões neutras e de raiva atingiu alto índice de acertos, ao contrário da tristeza que não foi detectada satisfatoriamente, pois o comportamento da boca foi erradamente interpretado pelo software, inclusive indicando estados de alegria confundindo gestos de sorriso com gestos de boca de preocupação. Uma dificuldade identificada foi a iluminação da sala que devido à natureza do sensor não reconhecia corretamente o usuário. A realimentação obtida por esse sistema permitiu através das unidades de animação, identificar em que estado as sobrancelhas e bocas foram seguidas para determinar a expressão pela malha.

No caso do Kinect 2.0, não se conta mais com uma malha triangular já que devido a maior resolução da câmera o modelo renderizado apresenta uma face de alta definição. Neste caso, a informação de profundidade permite detectar e apresentar a partir da malha de forma mais precisa as expressões, a Figura 3 apresenta um comparativo entre as imagens do sensor V1 e o sensor V2, delas, é possível observar os melhores resultados na geração do modelo 3D. Os resultados obtidos com o sensor V2 só apresentaram dificuldades da detecção dos movimentos da boca no cenário de dor onde os lábios são puxados para trás.

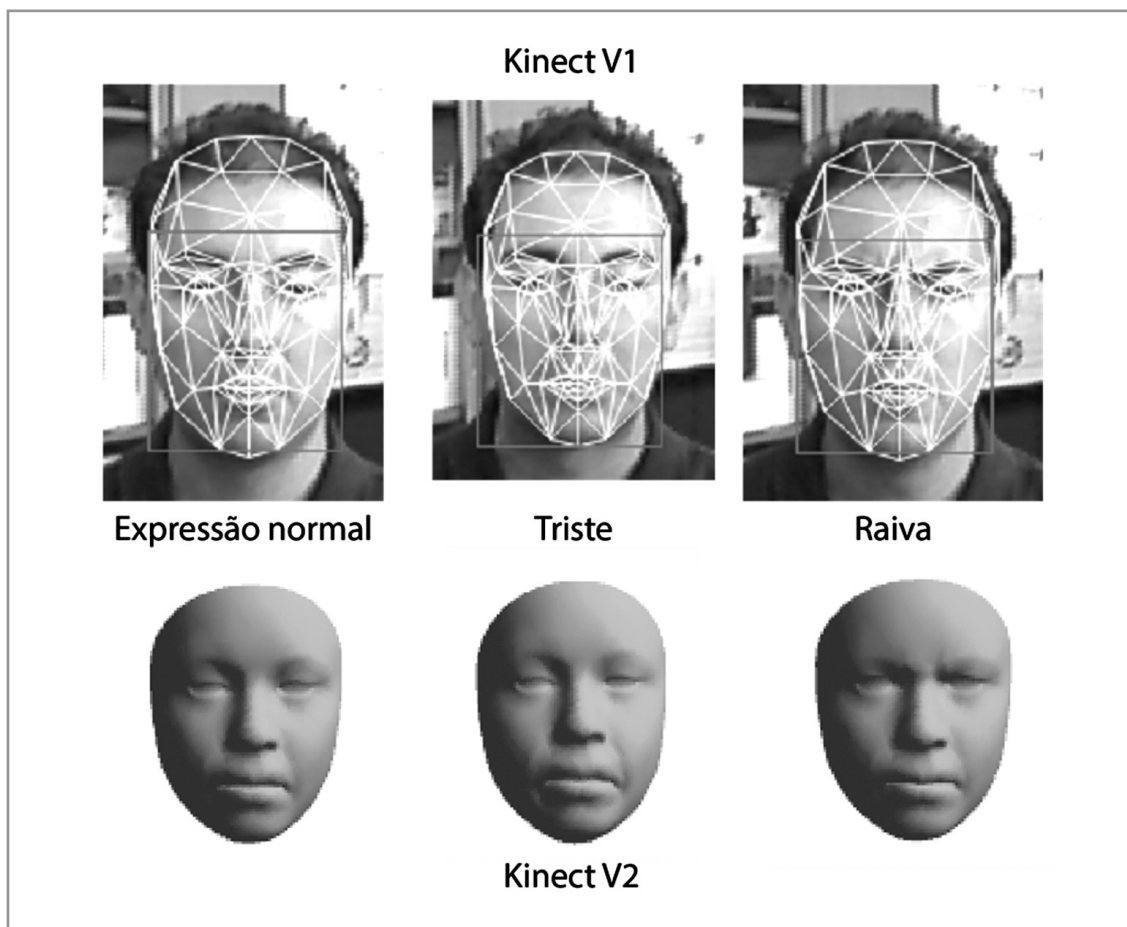


Figura 3. Expressões reconhecidas pelos Kinect V1 e V2
Fonte: Elaboração própria com base em informações.

Após realizar os testes surgiram dificuldades associados aos sensores como a incidência da luz na captura do rosto, o uso de óculos não fez diferença nenhuma e nenhum participante tinha barba para analisar se isto gerava alguma mudança no reconhecimento. A detecção dos estados triste apresentou dificuldades com o Kinect V1 já que os usuários manifestaram ter tido que exagerar os gestos para ser detectado, no caso do Kinect V2 esse problema não se apresentou.

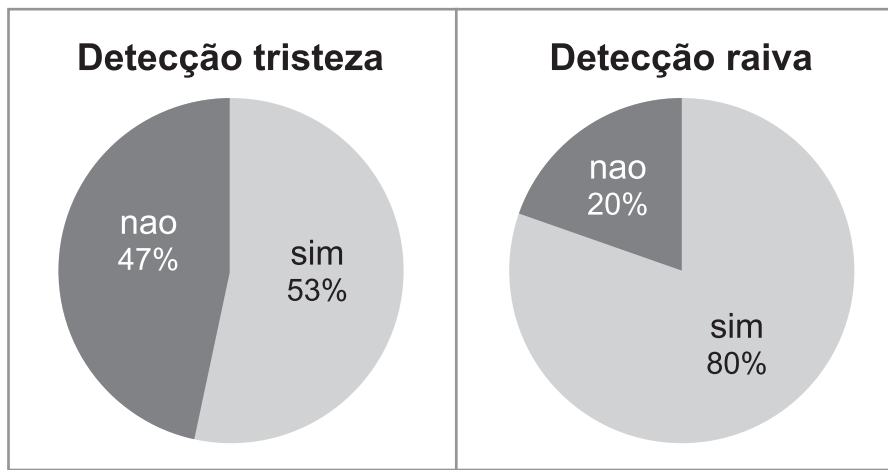


Figura 4. Porcentagem da detecção obtida
Fonte: Elaboração própria com base em informações.

Conclusões

Neste trabalho foi apresentado o desenvolvimento e implementação de um software para monitoramento de expressões do rosto comuns em situações de desconforto. Três expressões foram configuradas para a detecção, normal, triste e raiva, sua detecção foi validada através de testes com diferentes usuários com a intenção de verificar a adaptabilidade dos sensores nesses casos.

Os testes permitiram ratificar que a luz faz uma grande diferencia na captura do rosto sendo mais crítica com o Kinect V1 devido à resolução do vídeo processado, enquanto ambos sensores reconheceram os rostos, melhores resultados foram obtidos com o Kinect V2 graças a maior quantidade de vértices na malha da face, representando o comportamento do usuário. No caso dos resultados, a raiva foi melhor detectada devido aos movimentos das sobrancelhas cuja característica permite identificar o estado rapidamente, no caso da tristeza, devido aos movimentos musculares mínimos, os usuários tiveram problemas para expressá-la de forma que o sensor captara, neste cenário a exageração da expressão permitiu capturar melhor a face, mas resultou numa experiência para nada natural. Aplicações deste tipo de aplicativo podem encontrar uso em cenários de valoração psicológica a través de avatares virtuais que respondem ao usuário segundo seu rosto, voz e comportamento físico (DeVault, *et al.*, 2014), valoração de estresse como no caso dos pilotos de drones cujas atividades frente ao controle remoto e atividades põem em risco a saúde do usuário (Otto & Webber, 2013), cenários que podem ter impacto na avaliação e estudo de tomada de decisões e os efeitos sobre a pessoa.

Os trabalhos futuros deste trabalho serão orientados á implementação do sistema dentro de um cenário de trabalho onde a identificação de expressões do rosto permita valorar quando uma ação gera efeitos negativos no usuário.

Referencias

1. Amon, C. & Fuhrmann, F. (2014). Evaluation of the spatial resolution accuracy of the face tracking system for kinect for windows v1 and v2. *Proceedings of the 6th Congress of the Alps Adria Acoustics Association*.
2. DeVault, D., Artstein, R., Benn, G., Dey, T., Fast, E., Gainer, A. & Yu. (2014). SimSensei Kiosk: A virtual human interviewer for healthcare decision support. *Proceedings of the 2014 international conference on Autonomous agents and multi-agent systems*. (pp. 1061-1068).
3. Francese, R., Passero, I. & Tortora, G. (2012). Wiimote and Kinect: gestural user interfaces add a natural third dimension to HCI. *Proceedings of the International Working Conference on Advanced Visual Interfaces*. (pp. 116-123). ACM.
4. González-Ortega, D. D.-P.-Z., Antón-Rodríguez, M., Díez-Higuera, J. & Boto-Giralda, D. (2010). Real-time hands, face and facial features detection and tracking: Application to cognitive rehabilitation tests monitoring. *Journal of Network and Computer Applications*, 33(4), 447-466.
5. Lahajnar, F., Pernuš, F. & Kovačič, S. (2002). Machine vision system for inspecting electric plates. *Computers in industry*, 47(1), 113-122.
6. Matsumoto, Y., & Zelinsky, A. (2000). An algorithm for real-time stereo vision implementation of head pose and gaze direction measurement. Automatic Face and Gesture Recognition, 2000. Proceedings. *Fourth IEEE International Conference on* (pp. 499-504). IEEE.
7. Mehdi, Y. (Fevereiro de 2013). *Xbox Execs Talk Momentum and the Future of TV*. Fonte: *Microsoft news center*. From <http://news.microsoft.com/2013/02/11/xbox-execs-talk-momentum-and-the-future-of-tv/>
8. Otto, J. & Webber, B. (2013). Mental health diagnoses and counseling among pilots of remotely piloted aircraft in the United States Air Force. *Medical Surveillance Monthly Report*, 20(3), 3-8.
9. Paul, W. (2005). *Unmasking the Face: A Guide to Recognizing Emotions From Facial Expressions*. Los Altos, CA: Malor Books.
10. Perez, C., Palma, A., Holzmann, C. & Pena, C. (2001). Face and eye tracking algorithm based on digital image processing. *Systems, Man, and Cybernetics, 2001 IEEE International Conference on*. (pp. 1178-1183).
11. Samal, A. & Iyengar, P. (1992). Automatic recognition and analysis of human faces and facial expressions: A survey. *Pattern recognition*, 25(1), 65-77.
12. Sanz, J. & Jain, A. (1986). Machine-vision techniques for inspection of printed wiring boards and thick-film circuits. *JOSA*, 3(9), 1465-1482.
13. Schalkoff, R. (1989). *Digital image processing and computer vision*. New York: Wiley.
14. Share, N. (2015). *Desktop Operating System Market Share*. Fonte <http://www.netmarketshare.com/operating-system-market-share.aspx?qprid=10&qpcustomd=0>
15. Tsai, C. & Song, K. (2006). Face tracking interaction control of a nonholonomic mobile robot. *Intelligent Robots and Systems, 2006 IEEE/RSJ International Conference on*. (pp. 3319-3324).
16. Wang, S. (2006). Face-tracking as an augmented input in video games: enhancing presence, role-playing and control. *Proceedings of the SIGCHI conference on Human Factors in computing systems*. (pp. 1097-1106). ACM.
17. WHO. (2014). *Disability and health*. World Health Organization.
18. WHO. (Fevereiro de 2014). *eHealth*. Fonte: *World Health Organization*. From <http://www.who.int/topics/ehealth/en/>
19. Wingrave, C., Williamson, B., Varcholik, P., Rose, J., Miller, A., Charbonneau, E. & LaViola, J. (2010). The wiimote and beyond: Spatially convenient devices for 3d user interfaces. *Computer Graphics and Applications*, 30(2), 71-85.
20. Yang, X., Hasan, K., Bruce, N., & Irani, P. (2013). Surround-see: enabling peripheral vision on smartphones during active use. *Proceedings of the 26th annual ACM symposium on User interface software and technology*. (pp. 291-300).
21. Ye, Y., & Tsotsos, J. (1999). Sensor planning for 3D object search. *Computer Vision and Image Understanding*, 73(2), 145-168.
22. Yow, K. & Cipolla, R. (1997). Feature-based human face detection. *Image and vision computing*, 15(9), 713-735.
23. Zhang, Z. (2012). Microsoft kinect sensor and its effect. *MultiMedia*, 19(2), 4-10.