MEDIDA AUTOMÁTICA DO NÍVEL DE ATENÇÃO DE MOTORISTAS VIA VISÃO COMPUTACIONAL

joedlopes@gmail.com*, Thomas Brandmeier** e Alessandro Zimmer*

*Universidade Federal do Paraná, Curitiba, Brazil
** Technische Hochschule Ingolstadt, Ingolstadt, Alemanha
e-mail: joed@ufpr.br

Resumo: O presente artigo apresenta um modelo para estimar automaticamente - via visão computacional, a partir de imagens de vídeo capturadas por uma única câmera monocular (de baixo custo) com iluminação infravermelho, localizada no interior do veículo e de frente para o motorista - o nível de atenção relativa do condutor do veículo através de uma função projetada. Algoritmos para detecção da direção do olhar foram avaliados e adaptados. Além disso, um classificador baseado em redes neurais foi utilizado para detectar se os movimentos dos lábios são referentes à fala ou não. Imagens de 20 pessoas dirigindo um simulador, em um ambiente projetado, foram coletadas a fim de criar uma base de dados própria com cenários específicos. Para visualização, os resultados extraídos são reproduzidos em um modelo 3D, em um ambiente virtual desenvolvido.
Palavras-chave: Nível de Atenção de Motoristas, Visão Computacional, Detecção da Direção do Olhar, Detecção de Fala.

Abstract: *This paper presents an autonomous model to estimate – via computer vision, with images captured from only one monocular camera (low cost), in front of the driver inside the car – the attention's level by a designed function. Algorithms to detect the head pose and eye gaze were evaluated and adjusted. Furthermore, a classifier based on artificial neural networks was designed in order to classify the lip movements as talking or no word talking. Images from 20 volunteers were collected during a simulator driving to create a database with specific scenarios. For visualization and analysis of the extracted data by the model, a system to reproduce in a 3D model this was developed.*
Keywords: *Car Driver Attention Level, Computer Vision, Eye Gaze Detection, Speech Detection via Computer Vision.*

Introdução

A distração de motoristas de automóveis durante a condução é um dos principais causadores de acidentes de trânsito no Brasil, haja vista que os fatores característicos destes acidentes podem estar relacionados a todo o sistema de trânsito. A realização de mais tarefas, além da condução, como, por exemplo, conversar com passageiros ou usar telefone celular, bem como o uso de outros dispositivos eletrônicos no interior do automóvel ou ingerir gêneros alimentícios durante a condução, podem afetar diretamente o desempenho do motorista, diminuindo a sua atenção e outros processos cognitivos, tal como a capacidade motora e a tomada de decisões [1].

No âmbito da visão computacional (CV) modelos para monitoramento comportamental de motoristas podem ser utilizados justamente pelo fato de não serem intrusivos, dispensando o uso de sensores de movimentos, eletrodos e outros equipamentos intrusivos. Porém, fatores psicológicos e mentais do motorista não podem ser extraídos via CV, limitando-se apenas a detecção de movimentos do corpo do motorista.

Abordagens via CV para detecção da atenção visual do motorista tem sido propostas. Em [2] é apresentada uma abordagem para detecção da posição da cabeça do motorista, utilizando câmeras frontais e com suporte a iluminação infravermelho (IR), permitindo a captura de imagens em ambientes com pouca iluminação.

Em [3] foi desenvolvido um sistema que utiliza imagens capturadas por câmeras 3D no interior do veículo, possibilitando a detecção de movimentos acima da cintura do motorista. Outra câmera, monocular (2D), no interior do veículo é utilizada para detecção da direção do olhar. A desvantagem deste modelo é o custo da câmera 3D, pois é relativamente mais caro se comparado com o valor das câmeras monoculares.

Um sistema para alerta de colisões frontais baseado no foco visual é proposto em [4], tal abordagem utiliza imagens monoculares frontais e iluminação IR. Para estimar o foco visual foi criado um classificador, que a partir dos ângulos da posição da cabeça é determinado para qual região no interior do veículo o motorista está olhando. Esta abordagem utiliza somente a posição da cabeça para determinar o foco visual, não relacionando a posição da íris em relação ao contorno dos olhos.

Em [2] e [4] foram desenvolvidas bases de dados utilizando sensores inerciais (captura de movimentos da cabeça), os quais fornecem uma informação exata da direção da cabeça, sendo utilizados na etapa de validação dos métodos propostos. No que se refere ao ambiente para coleta de dados, especificamente ao veículo, podem ser utilizados simuladores, nos quais são possíveis submeter os indivíduos a situações e eventos específicos, sem comprometer a integridade física destes.

Dentre as abordagens citadas, foi utilizado apenas o foco visual, não relacionando outros fatores, como a detecção de fala.

O objetivo deste projeto é propor um modelo para, a partir da detecção de fala e do foco visual via visão

computacional, medir automaticamente o nível de atenção relativa do motorista durante a condução. A única informação de entrada são imagens monoculares, providas por uma única câmera com iluminação IR localizada no interior do veículo de frente para o motorista. Uma função iterativa para calcular o nível de atenção com base nas informações detectadas a cada quadro foi projetada. O nível de atenção medido pode ser utilizado para aumentar a segurança da condução, possibilitando a ativação prematura dos dispositivos de segurança passivos do veículo com o intuito de minimizar danos aos condutores, passageiros e terceiros. Outro objetivo do projeto foi criar uma base de dados de motoristas dirigindo um simulador, submetendo-os a situações e cenários específicos projetados previamente.

Materiais e métodos

Antes de começar a coleta de dados, é necessário escolher uma câmera que seja compatível com todos as características do ambiente. A câmera escolhida foi uma *GoPro HD Hero*, cuja resolução é de 1920x1080 pixels. As lentes originais foram substituídas por lentes de 2.8 milímetros e o filtro infravermelho removido, permitindo assim capturar imagens do motorista e do passageiro ao seu lado. Além disso, foi utilizada uma luminária com luzes IR, com frequências entre 780 e 880 nanômetros, lembrando que tal espectro não é visível ao olho humano e não interfere na condução do motorista. As posições da câmera e da luminária utilizada são mostradas na Figura 1 e ambas distam, em média, 61 cm da cabeça do motorista, além de estarem a baixo da linha de visão. Portanto, não dificultam a visão e, consequentemente, não causam distrações e dispensam a realização de normalizações provenientes da projeção.

Figura 1 – Interior (esq.) e exterior (dir.) do simulador.

A coleta de dados foi realizada na *Technische Hochschule Ingolstadt*, na Alemanha, a qual disponibilizou um simulador realístico, denominado HEXAPOD, capaz de realizar movimentos com até 6 graus de liberdade, sendo constituído pela parte frontal de um Volkswagen Golf, conforme mostra a Figura 1. Um computador é o responsável por controlar todos os sensores e atuadores do veículo e por meio do software IPG CarMaker é possível projetar todo o sistema viário, trafego e também capturar as informações do veículo e as ações do motorista durante a simulação.

Com a colaboração de integrantes do departamento de psicologia da Universidade Federal do Paraná, foram elaborados cenários e experimentos baseados na realidade do trânsito brasileiro. Vale destacar que os problemas relativos à condução foram devidamente recriados no ambiente virtual, fato este que propicia uma maior veracidade na obtenção dos resultados.

Ao todo foram criadas duas bases de dados. A primeira, de maior relevância para o projeto, baseia-se na condução do veículo no trecho da BR-277, rodovia que interliga a cidade de Curitiba à cidade de Paranaguá, contendo várias circunstâncias que possivelmente podem configurar distrações para o motorista. Em determinado momento da simulação foi solicitado ao motorista que conversasse com um passageiro e, em outro instante, que atendesse um telefone celular.

A segunda base é utilizada para calibração da detecção do foco visual, na qual o motorista visualiza, por 10 segundos, determinadas regiões do interior do simulador como, por exemplo, rádio, velocímetro, espelhos, porta luvas, freio de mão e também para frente.

Tais experimentos propostos foram realizados com 20 voluntários, todos estudantes universitários com idade entre 20 e 31 anos, dentre eles, 10 brasileiros, 5 alemães, 2 mexicanos, 1 indiano, 1 paquistanês e 1 vietnamita. A escolha dos participantes foi realizada de forma aleatória, sendo que nenhum dos indivíduos obteve conhecimento prévio do experimento. Além disso, fora solicitado o preenchimento de um formulário com questões que versavam sobre trânsito e que assinassem um termo, cuja finalidade era a disponibilização do uso das informações coletadas. Na Alemanha, para este tipo de coleta de dados não é necessário a aprovação do comitê de ética, apenas a assinatura de um termo de consentimento pelos voluntários.

Durante todo o processo de coleta de dados, foi utilizado um sensor de captura de movimentos localizado na cabeça do motorista (ver Figura 1), para criar um conjunto verdade para validação do método de detecção da posição da cabeça durante a condução.

Conforme mencionado, a única informação de entrada do sistema são imagens monoculares e, portanto, foram utilizados apenas métodos de visão computacional baseados em imagens monoculares.

A Figura 2 apresenta um diagrama simplificado do modelo proposto, o qual pode ser dividido em três etapas: captura, detecção e fusão. A captura baseia-se em ler arquivos de vídeo provenientes da coleta de dados e fornecer sequencialmente as imagens para os métodos posteriores. A detecção é composta por algoritmos, que visam detectar a direção do olhar e fala do motorista. Por fim, a fusão de tais informações extraídas do quadro atual é utilizada para calcular o nível de atenção.

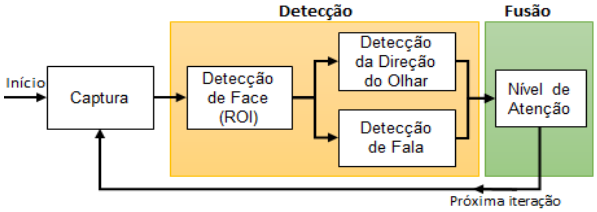

Figura 2: Diagrama simplificado do modelo proposto.

A partir da imagem de entrada é detectada uma região de interesse (ROI), isto é, a face do motorista, haja vista

que os métodos posteriores utilizam apenas características faciais. Para detecção inicial da ROI e a cada 30 quadros, é aplicado o método proposto em [5] na imagem de entrada com tamanho reduzido a 10% do tamanho original e em tons de cinza, com o objetivo de diminuir o custo computacional do método. Durante o intervalo de 30 quadros, é utilizado um método desenvolvido para rastrear a face do motorista, que consiste no cálculo do fluxo óptico [6] para uma grade com 49 pontos situados na ROI. Através do movimento médio destes pontos, a posição da ROI é atualizada. Embora utilize imagens em menor escala, o método de detecção de face utilizado é computacionalmente mais complexo do que o método de rastreamento.

Abordagens para detecção da posição da cabeça foram analisadas e optou-se por utilizar um modelo deformável da face, conforme proposto em [7] e [8]. Tal modelo baseia-se na detecção individual dos 66 pontos característicos e através de métodos de otimização (mínimos quadrados) é encontrado o melhor ajuste. Além deste modelo se ajustar às características faciais, como, por exemplo, às variações da boca e olhos, pode-se obter também a rotação da face, com 3 graus de liberdade (*roll, pitch e yaw*).

Para calcular o foco visual, além de utilizar a posição da cabeça, este projeto busca relacionar a posição da íris ou pupila em relação aos contornos dos olhos, a fim de fornecer um valor mais acurado da direção do olhar. Para detecção da íris foi utilizado o algoritmo proposto em [9], porém no pré-processamento foi utilizado o algoritmo MeanShift [10], com a finalidade de remover os reflexos provenientes da iluminação de tal região dos olhos. Conforme a figura 3, observa-se que a partir da posição da íris encontrada, $P(x,y)$, um vetor diretor pode ser calculado com base nos ângulos α e β, onde x e y são normalizados em relação a w e os movimentos da íris podem ser realizados até um ângulo limite θ, definido em 30°, e por Φ calculado com base no valor de w.

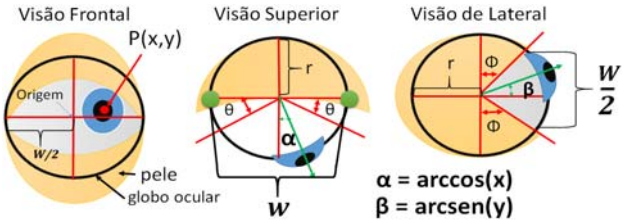

Figura 3 – Globo ocular e funções propostas

Os vetores diretores provenientes da posição da íris e posição da cabeça são somados (soma vetorial) de modo que resultam no vetor da direção do olhar ($\vec{v}$), com origem na localização da ROI (l) e com a informação de profundidade, isto é, o quanto o motorista está distante da câmera, calculado a partir da distância entre as pupilas (para imagens em que a face não está rotacionada).

Para obter o foco visual do motorista, regiões (planos formados por 2 triângulos retângulos), no interior do veículo foram definidas conforme a Figura 4, de forma que estão na mesma escala de l, possibilitando o cálculo da intersecção entre $\vec{v}$, situado na origem l, com triângulos (das regiões) através do algoritmo de intersecção entre vetores e triângulos proposto em [11].

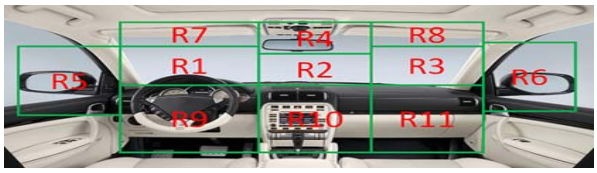

Figura 4 – Regiões definidas no interior do veículo.

Para a detecção de fala, foram estudadas em redes neurais [12] para projetar uma *Rede Neural Artificial Perceptron Multicamadas*, com 100 neurônios na camada oculta e com método de treinamento de Levenberg-Marquardt. O conjunto de entrada é um vetor de características contendo as variações das distâncias (ver Figura 5) *X2, X3, X4, X5*, e divididas por *X1* (tamanho horizontal da face) para normalização, ao longo dos ultimas 30 quadros, totalizando um vetor 120 características de entrada.

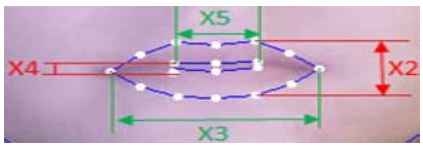

Figura 5: distância entre pontos da boca.

A figura 8 mostra o algoritmo para o cálculo iterativo do nível de atenção proposto, podendo variar de -1.0, situação em que o motorista está totalmente desfocado, a +1.0, situação em que o motorista está totalmente focado no trânsito. Cada região no interior do veículo tem um peso específico: R1 e R2 tem peso +0.2, R4, R5, R6, tem peso -0.1, R7 tem peso +0.05 e demais regiões -0,3. Caso o método de detecção de face ou foco visual falhe, serão atribuídos o valor negativo -0.03 para a iteração atual. E a ação de conversar tem peso -0,03.

Seja:

Vetor de regiões $\mathbf{R} = \{r_i, r_{i+1}, r_{i+2}, ..., r_n\}$,
1 quando ativada, 0 quando desativada

Vetor de pesos $\mathbf{W} = \{w_i, w_{i+1}, w_{i+2}, ..., w_n\}$
referente ao peso de cada região

C = motorista conversando (1), e não conversando (0)

Wc = peso referente a tarefa de conversar

A é o percentual de atenção.

Início:

1. A = 0
2. Enquanto Captura de Imagens Ativa
   - 2.1 Detectar Direção do Olhar
     - 2.1.1 Posição da Face
     - 2.1.2 Posição da Íris
   - 2.2 Calcular Intersecções
   - 2.3 Detectar Movimentos da Boca
   - 2.4 $A = A + (\sum_{i=1}^{n}(r_i \,.\, wi)) + C \,.\, Wc$
   - 2.5 Se A > 1.0, então A = +1.0
   - 2.6 Se A < -1, então A = -1.0

Figura 8: Algoritmo para cálculo do nível de atenção.

Para visualização dos dados extraídos foi desenvolvido um ambiente virtual 3D utilizando o motor gráfico Ogre3D.

Resultados

A base de dados para treinamento da detecção de fala, foi marcada de forma manual. Movimentos dos lábios de 3 sujeitos foram suficientes para criação de toda a base. A taxa de acerto, durante a etapa de validação para a detecção de fala foi de 82,3%.

A detecção da posição da cabeça teve uma taxa de acerto 85,23% na base de dados de validação, para ângulos de rotação de até ±60° nos três eixos, tal limite é suficiente para o presente projeto, haja vista que o espelho direito está localizado em média 45° a direita.

O sistema proposto foi utilizado para medir o nível de atenção dos brasileiros durante o experimento realizado no HEXAPOD e a Tabela 1 apresenta o valor médio do nível de atenção medido durante as distrações para todos os sujeitos brasileiros.

Tabela 1: Nível de atenção durante determinadas tarefas.

Distração	Nível de Atenção Médio	Tempo de Duração
1. Conversar com Passageiro	0.60 (±0.05)	2 minutos
2. Atender o Celular	-0.32 (±0.10)	15 segundos
3. Conversar usando o Celular	0.62 (±0.06)	2 minutos

Discussão

É possível observar a partir dos valores medidos (Tabela 1), que conversar com passageiros e conversar usando o celular tem quase a mesma importância, resultado semelhante foi obtido em [13], no qual verificou-se que não há diferenças significativas na condução entre motoristas segurando o celular ou usando fones de ouvido para conversar. Em [1] é mostrado uma diminuição na ativação do cérebro nas tarefas relacionadas a condução quando o motorista escutando alguém falar, tal fator pode comprometer a condução.

O nível de atenção durante a distração ‘atender o celular’ foi o que apresentou menor nível de atenção medido durante a condução, devido ao fato do motorista ter que olhar para regiões com pesos negativos.

A detecção do foco visual proposto neste artigo difere dos modelos propostos em [2], [3] e [4], que utilizam métodos de aprendizagem de máquina para detectar tal valor. O modelo proposto utiliza apenas os vetores diretores e intersecção com as regiões previamente definidas e também relaciona a posição da cabeça em três dimensões. Além disso, apresenta uma função para medir o nível de atenção, agregando ainda a detecção de fala.

Conclusão

Este projeto propôs um modelo para detecção do nível de atenção de motoristas. Porém, seria interessante agregar mais informações como, por exemplo, do veículo (velocidade, ângulo do volante, pedais, etc.), o padrão de direção de cada motorista e o reconhecimento de expressões faciais, a fim de aumentar a precisão do modelo proposto.

Agradecimentos

Os autores agradecem ao apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) e ao departamento de Engenharia Elétrica da Technische Hochschule Ingolstadt, por disponibilizar a estrutura necessária para coleta de dados e aos membros do departamento de Psicologia da Universidade Federal do Paraná envolvidos.

Referências

[1] Just MA, Keller TA, Cynkar JA. A decrease in brain activation associated with driving when listening to someone speak. Brain Research 1205. 2008; p. 70-80.

[2] Murphy-Chutorian E, Doshi A, Trivedi MM. Head Pose Estimation for Driver Assistance Systems: A Robust Algorithm and Experimental Evaluation. Intelligent Transportation Systems Conference, IEEE, 2007; p. 709-714.

[3] Guranaratne P, Mannino M. A closed-loop crash warning system based on heterogeneous cues. Automation, Robotics and Appliacations (ICARA). 2011; p. 63-66.

[4] Lee SJ, Jo J, Jung HG, Park KR, Kim J. Real Time Gaze Estimator Based on Driver’s Head Orientation for Forward Collision Warning System. Em: Trans. Intell. Transport. Sys. 12. 2011; p. 254-267.

[5] Viola P, Jones MJ. Robust Real-Time Face Detection. Int. J. Comput. Vision 57. May 2004; p. 137-154.

[6] Bouguet JY. Pyramidal Implementation of the Lucas Kanade Feature Tracker Description of the algorithm: Available from: http://robots.stanford.edu/cs223b04/algo_tracking.pdf

[7] Saragih J, Lucey S, Cohn J. Deformable Model Fitting by Regularized Landmark Mean-Shifts. International Journal of Computer Vision (IJCV), 2010.

[8] McDonald K. FaceTracker Library. Available from: https://github.com/kylemcdonald/FaceTracker

[9] Ibanez VBL, Yano V, Zimmer A. Automatic pupil size measurement based on region growth. Biosignals and Biorobotics Conference (BRC). 2012; p. 9-11.

[10] Tao W, Jin H, Zhang Y. Color Image Segmentation based on Mean Shift and Normalized Cuts. Systems, Man, and Cibernetics. Em: Cybernetics, IEEE Transactions on. vol 37. no. 5. 2007; p. 1382-1389.

[11] Möller T, Trumbore B. Fast, minimum storage ray-triangle intersection. J. Graph. Tools 2. 1997; p. 21-28.

[12] Waibel A. Modular construction of time-delay neural networks for speech recognition. Neural Comput. 1. 1989; p. 39-46.

[13] Strayer DL, Drews FA, Crouch DJ. A comparison of cell phone driver and drunk driver. HF. 2006;