Automatic Detection of Human Faces in Uncontrolled Environments:

identification of direction and movement

Miguel Coelho de Pinho, Nuno Magalhães Ribeiro, Feliz Ribeiro Gouveia
(15526@ufp.edu.pt, nribeiro@ufp.edu.pt, fribeiro@ufp.edu.pt)
CEREM – Centro de Estudos e Recursos Multimediáticos da Faculdade de Ciência e Tecnologia
Universidade Fernando Pessoa, UFP, Porto, Portugal

Abstract **— This paper presents an application for automatic face detection on video streams from surveillance cameras in public or commercial places. In many situations it is useful to detect where to are people looking for, e.g. in exhibits, commercial mals, and public places in buildings. Our application is designed to work with surveillance cameras that are already available in those places, and do not imply an extra acquisition cost. The paper begins with a review of techniques used for face detection, a brief introduction to the OpenCV library, and of the requirements for the application. We then describe our approach, and present the main algorithms we used. We then perform an evaluation of the application, analyzing its performance throughout the processing, and analyze the accuracy of the recognition. We performed tests in real-time and off-line. The results are encouraging, and we identify limitations and improvements that can be introduced to decrease the error rate.**

Keywords: Multimedia Processing and Analysis; Processing and Analysis of Digital Video; Automatic Detection of Human Faces; OpenCV; Haar Filters.

I. INTRODUÇÃO

A detecção e o seguimento de faces humanas em sequências de vídeo é uma área muito importante e fundamental na visão computacional[1].É uma das áreas especialmente activa devido às suas inúmeras aplicações e à complexidade a elas inerente. De facto, não existem ainda soluções suficientemente desenvolvidas que permitam que as empresas processem os registos vídeo obtidos através de câmaras de videovigilância de forma a auxiliar a tomada de decisão sobre a actividade humana nas áreas e locais abrangidos pela câmara. Neste contexto, o desenvolvimento deste trabalho apresenta uma solução para este problema, consistindo em duas aplicações de análise de trajectos e orientação de pessoas em ambientes não controlados.

A inexistência de meios automáticos e de baixo custo para se obter informação sobre as orientações do olhar das pessoas e sobre as respectivas deslocações num determinado espaço físico, obriga os profissionais a observar as gravações, ou as imagens captadas directamente pela câmara. Esta forma de proceder conduz a um grande desperdício de tempo e a informação pouco fidedigna devido ao cansaço que produz nas pessoas. Assim, a motivação para este trabalho foi a de tentar reconhecer automaticamente faces humanas e determinar a sua direcção com uma simples câmara de vídeo para que possa ser utilizada em ambientes não-controlados sem introduzir grandes custos. Os cenários de utilização possível são as grandes superfícies comerciais tais como os hipermercados, lojas temáticas, áreas de exposições, estações de comboios ou de transportes terrestres, aeroportos e outras localizações onde seja importante identificar o que prende a atenção das pessoas. Nestes contextos, é útil efectuar o processamento e a análise de imagens (*frames* de vídeo) obtidas a partir de câmaras de videovigilância já instaladas, não sendo necessário recorrer a equipamento mais sofisticado, ou sequer a alterações de equipamentos existentes, limitando assim os custos. Este artigo encontra-se organizado em seis secções. A seguir esta introdução, na secção 2 efectua-se uma revisão sucinta do estado da arte das aplicações de detecção facial. A secção3 descreve a forma como foi abordado o problema. É ainda apresentada a arquitectura das aplicações, designadas por "FaceDetectRT" e "FaceDetectVC". A secção 4 descreve a avaliação que foi realizada sobre as aplicações desenvolvidas. Na última secção são apresentadas as conclusões do estudo descrito neste artigo, as suas limitações e o trabalho futuro.

II. DETECÇÃO DE FACES HUMANAS

Segundo Benezethetal.[1], o interesse da criação de aplicações que detectam humanos tem vindo a aumentar nos últimos anos devido ao aumento da robustez, entre outras, nas áreas da videovigilância, na robótica, e na indexação baseada em conteúdos multimédia. De facto, Kollreideretal.[2] afirmam que a análise facial biométrica pode ser baseada em duas aproximações: (1) Detecção facial e (2) Reconhecimento facial. A detecção facial permite localizar uma ou mais instâncias correspondentes à face humana contidas num vídeo ou numa imagem. Por outro lado, o reconhecimento facial estabelece uma relação única entre dois ou mais traços independentes que correspondem ao rosto dessa mesma pessoa[2]. Os desafios, tanto em relação à detecção humana como à detecção facial, dizem respeito ao ambiente onde essas imagens são retiradas, pois a complexidade do ambiente dificulta a acção da detecção devido a aspectos tais como, entre outros, a iluminação, plano de fundo, cor da pele, e expressões faciais, [1]e[2]. Já a técnica proposta por Viola & Jones[3] dá a possibilidade de efectuar a detecção de qualquer tipo de objecto (neste caso faces humanas) com uma forma genérica e em tempo real com uma taxa de sucesso bastante elevada. A abordagem proposta baseia-se em algoritmos de aprendizagem automática (*machinelearning*) aplicados de forma a identificar objectos, e capazes de processar imagens com muita rapidez. A utilização das características de um objecto, ao invés dados pixéis

contidos numa imagem, permite aumentar a velocidade da análise porque o número de características de um objecto é substancialmente inferior ao número de pixéis de um objecto[3]. Na concepção da sua ferramenta de detecção, os autores principiaram por basear-se na técnica proposta por Papageorgiouetal.[4] na qual a representação do objecto é feita de forma rectangular dentro de uma área de detecção baseada em cálculos semelhantes aos de uma transformada de Haar, designados por *Haar-like* (ver figura 1).

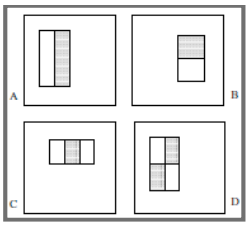

Figura 1. Características Rectângulares

Na figura1 ilustram-se três tipos distintos de características: (1) a característica de dois rectângulos (A e B) cujo valor é a diferença entre a soma dos pixéis numa área com dois objectos rectangulares do mesmo tamanho; (2) a característica de três rectângulos (C) calcula a soma dos dois rectângulos exteriores e subtrai com o do centro; e (3) a característica de quatro rectângulos (D) que faz a diferença entre os pares das diagonais dos rectângulos. De forma a aumentar a rapidez da computação destas características rectangulares, Viola & Jones[3] usaram uma representação intermédia que designaram por imagem integral ("*Integral Image*"). Esta define que numa dada localização, definida por um par de coordenadas *x* e *y*, é efectuado o somatório de todos os pixéis acima e à esquerda do ponto definido. Após obter os cálculos de todos os pixéis nessa área é possível obter os cálculos de qualquer área rectangular através de uma representação matricial de quatro referências, ilustrada na figura 2, onde é possível determinar a região D sabendo que o valor na posição 1 é a soma dos pixéis do rectângulo A, o valor da posição 2 será A+B, a posição 3 será A+C e a posição 4 será A+B+C. Tendo as quatro referências, a soma de D é definida pelos valores das posições 4+1 – (2+3).

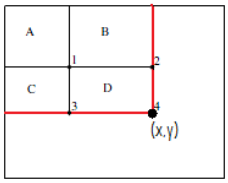

Figura 2. Matriz de quatro Referências

O passo seguinte consiste em treinar os classificadores através de um conjunto de imagens no qual existem imagens positivas e negativas: as positivas contêm o objecto que se deseja encontrar e as negativas evidenciam a ausência desse objecto. Os autores utilizaram uma variante do algoritmo "*AdaBoost*" criado por Freund[5] para seleccionar um pequeno conjunto de características e treinar o classificador. De forma a descartar várias áreas da imagem onde não existe o objecto desejado, os autores utilizaram classificadores em cascata[3] possibilitando assim um maior foco nas áreas mais promissoras da imagem. A cascata de classificadores proposta possui *n* etapas, sendo que em cada etapa se determina se a área de pesquisa pode ou não corresponder a uma face humana. Caso o objecto seja aceite como possível face, passa então para a etapa seguinte; caso contrário será rejeitado. Em cada etapa, os classificadores são cada vez mais precisos. Caso a área em análise na última etapa seja classificada como face, então assegura-se, com uma probabilidade entre 85 e 95%, que nessa área existe o rosto de uma pessoa. No sentido de melhorar este método, Lienhart&Maydt[6] decidiram expandir as representações quadráticas *Haar* através da introdução de novas representações reduzindo a taxa de falsos positivos em cerca de 10% em relação ao método anterior, possibilitando pois um aumento na eficácia da detecção de faces. Já a aplicação proposta por Benezethetal.[1] pretendeu obter um sistema com um elevado grau de desempenho para ser usado em ambientes interiores não controlados. O algoritmo que propuseram para a realização da detecção em tempo real do corpo de um ser humano em vídeos baseia-se em 3 fases distintas. Inicialmente recorreram à técnica de subtracção do plano de fundo (*background subtraction*) que permite limitar o espaço de pesquisa dos classificadores *Haar-like* retirando as áreas de interesse que se encontram na imagem. A escolha destas áreas faz-se através do cálculo da diferença entre a imagem actual e a imagem anterior, permitindo encontrar as movimentações dos objectos presentes na imagem. O passo seguinte corresponde à classificação de cada objecto retirado através da subtracção do fundo (designado por "*blob*"). Cada um destes objectos será posteriormente seguido individualmente, já que pode corresponder a uma pessoa. Finalmente utilizam-se os classificadores *Haar-like* para determinar de forma rápida se na verdade se trata de uma pessoa e, portanto, aceitar ou rejeitar o objecto. Os resultados obtidos por Benezethetal.[1] foram de aproximadamente 97% em ambientes interiores não controlados para os quais os autores afirmam que o método de subtracção permite reduzir o número de falsas detecções via redução do espaço de pesquisa dos classificadores.

III. FERRAMENTAS FACEDETECT

No sentido de conceber uma possível solução para o problema introduzido acima, pretendeu-se tirar partido das tecnologias existentes. Os resultados obtidos permitem afirmar que a adição de novas funcionalidades a tais tecnologias permite melhorá-las em aspectos menos explorados até ao momento. Para tal, foram desenvolvidas duas aplicações:

- A aplicação *FaceDetectRT* que permite detectar faces humanas e outros elementos de cada face (olhos, nariz) de uma pessoa em vídeos adquiridos em tempo real.
- A aplicação *FaceDetectVC* que permite efectuar a detecção e seguimento de rostos humanos em ficheiros de vídeo já existentes, proporcionando assim maior liberdade de utilização em diferentes situações e contextos.

De forma a cumprir os objectivos propostos, apresenta-se a seguir todos os requisitos funcionais e não-funcionais. Em termos dos requisitos funcionais, a aplicação que efectuará a detecção de faces em ficheiros de vídeo existentes deverá (1) permitir ao utilizador o processamento de qualquer vídeo desejado. Ambas as aplicações devem (2) permitir ao utilizador analisar a informação recolhida de uma face à sua escolha após

o processamento de um vídeo ou de uma recolha em tempo real. Esta informação será composta pelo trajecto da face detectada que pertence à pessoa em questão, e essencialmente, identificar-se-ão os lados para os quais a pessoa olhou (esquerda ou direita), bem como o tempo passado em cada uma dessas orientações. Para além disso, as aplicações devem (3) indicar claramente quantas orientações ocorreram no sentido da direita ou da esquerda em todas as detecções faciais analisadas.

Em relação aos requisitos não-funcionais, as aplicações devem (1) ser simples de utilizar e devem exigir um grau de aprendizagem tão reduzido quanto possível. O menu principal deve ser facilmente acedido após o processamento das *frames* de vídeo. Ainda (2) a detecção de faces, bem como os respectivos elementos característicos, os olhos e o nariz, devem ser identificados automaticamente sem a intervenção do utilizador. Por outro lado, (3) as informações relevantes recolhidas durante o processamento devem ser visualizadas pelo utilizador. Pretende-se ainda que (4) as aplicações possibilitam um processamento rápido de *frames* para que a visualização em tempo real não seja comprometida. Após a revisão bibliográfica efectuada sobre as técnicas específicas que têm sido utilizadas para o reconhecimento de faces humanas, incluindo, entre outras, a subtracção do fundo e filtros com informações da cor da pele, o método escolhido para a resolução do problema descrito foi o da utilização de classificadores *Haar* em cascata proposto em[3], com os melhoramentos propostos em[6]. Esta escolha deveu-se ao facto destes classificadores realizarem a detecção nos dados após estes serem disponibilizados pelo utilizador sob a forma de um ficheiro de vídeo, ou directamente na *stream* de vídeo proporcionada por uma câmara de vídeo, permitindo assim a possibilidade de efectuar detecções em tempo real para além do bom desempenho, a eficiência da detecção e o baixo custo de computação[1]. Finalmente, os classificadores de *Haar* podem ainda ser treinados para efectuar detecções dos objectos através de um conjunto de amostras positivas e negativas.

A. *Estrutura de Armazenamento dos Dados*

A estrutura de dados seleccionada para o armazenamento da informação recolhida foi a lista, pois, para além de permitir uma organização dos dados de forma encadeada, possui também os requisitos necessários para as aplicações, já que proporciona uma forma rápida e eficaz de armazenar temporariamente os dados em memória. A figura 3 ilustra a estrutura das listas ligadas desenvolvida para este efeito. Note-se que cada nó da Lista de Faces possui informações relativas à face detectada de cada pessoa. A segunda lista contém todas as coordenadas cartesianas (x,y) adquiridas nas detecções de faces, olhos e narizes de cada pessoa, o que permite obter informação sobre os percursos das pessoas no espaço físico observado, bem como as respectivas orientações, e permitem ainda controlar o processo da detecção na medida em que identificam as faces de pessoas que já foram detectadas em frames anteriores do vídeo que se está a processar. Em termos de informação contida em cada nó existente na Lista de Faces descritas acima, atente-se à descrição apresentada na tabela 1.

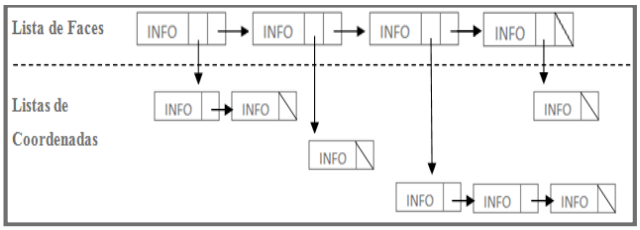

Figura 3. Estrutura das Listas Ligadas

Tabela I. LISTA DE VARIÁVEIS/ INFORMAÇÕES DOS NÓS

Nó da Lista de:	Nome	Descrição
Faces	ID	Valor que permite identificar uma face de forma única.
	NºFrames	Representa o número de *frames*de vídeos que passaram desde a primeira detecção de uma determinada face.
	NºPontos	Número total de pontos guardados na lista de coordenadas desde a primeira detecção de uma determinada face.
	Fim	Variável binária que permite determinar a finalização de detecções referentes a uma determinada face.
	NóPontos	Elo de ligação à lista de coordenadas.
	Elo	Elo de ligação às faces seguintes.

B. *Implementação das aplicações*

Descrevemos agora a implementação das aplicações *FaceDetectVC* e *FaceDetectRT* detalhando o trabalho em cada uma das seguintes fases: análise de *frames*, visualização de *frames*, eliminação de faces e apresentação do menu de resultados adquiridos. Ambas foram implementadas recorrendo à ferramenta de desenvolvimento *Visual Studio 2008* e à biblioteca *OpenCV v.2.0*da Intel [7].

1) *Análise de Frames*

Cada *frame* recebida por recolha directa de uma câmara de vídeo, ou adquirida a partir de um ficheiro de vídeo, é analisada de forma a identificar a face, os olhos e o nariz de uma pessoa. Cada identificação será assinalada através da inscrição de: (i) um rectângulo sobre a *frame* de vídeo no caso da face, (ii) um círculo para identificar a localização de ambos os olhos, e (iii) uma linha recta desde o centro da margem superior do rectângulo até ao centro do nariz para identificar a orientação da face. Ainda nesta fase, faz-se a inserção dos dados recolhidos na lista ligada de acordo com a informação que é adquirida durante o reconhecimento, determinando-se igualmente a orientação da face. Inicialmente foi necessário associar variáveis para cada tipo de ficheiro de classificadores *Haar* a utilizar. Estes classificadores constituem a base principal da detecção facial pois cada um deles efectua os cálculos que foram definidos no método sugerido em[3]. Os dados contidos nos classificadores são representações da "imagem integral" de amostras positivas e negativas de um objecto a detectar. Neste caso, os objectos de interesse são as faces, os narizes e os olhos das pessoas. Sabendo da importância da obtenção de um classificador bem construído, optou-se por utilizar classificadores *Haar* que são fornecidos pela Intel através da instalação da biblioteca *OpenCV*. No decorrer da captura, cada *frame* é analisada de forma individual e reescrita no sentido de indicar as detecções que foram

realizadas. Nesta fase de análise, as aplicações começam por criar uma cópia da *frame* original com um tamanho reduzido, sendo as cores convertidas para uma escala de tons de cinzento após aplicar o grau de prioridade BGR. De forma a aumentar a eficácia da detecção, utilizou-se uma função para equalizar o histograma que permite normalizar as escalas de brilho. Dado que o método proposto em[3] possui informações de brilho acerca de um determinado objecto a detectar, isto é necessário para ajudar a reconhecer com mais facilidade esse objecto. A necessidade de redimensionamento da imagem proporciona um aumento do desempenho das detecções devido ao facto de os classificadores terem deste modo menos áreas a percorrer do que na imagem original.

a) Detecções Faciais/Oculares/Nasais

Tomadas as medidas descritas acima, inicia-se a pesquisa de faces através da função "cvHaarDetectObjects()" que aplica o método descrito em[3]. Esta função permite descobrir áreas rectangulares na imagem que são idênticas aos objectos guardados nos classificadores *Haar*. Assim, compara a área da imagem com cada etapa dos classificadores; caso conclua todas as etapas, e se os objectos dos classificadores corresponderem a faces humanas, então o objecto analisado tem grande probabilidade de ser uma face humana. A análise da imagem é efectuada várias vezes com diferentes escalas. Em casos de sobreposição de imagem, tais como a análise de áreas onde já foi efectuada uma análise pelo classificador, a função aplica heurísticas de forma a reduzir o número de regiões analisadas. No final da pesquisa, a função retorna uma sequência de rectângulos que são susceptíveis de conter o objecto para o qual o classificador está treinado. De modo a que o utilizador possa visualizar as faces encontradas, as aplicações desenvolvidas utilizam as coordenadas dos rectângulos reconhecidos pela função anterior para desenhar na imagem original o rectângulo que representa a face. Estas coordenadas são necessárias não só para a representação da face, mas também para diminuir a área de procura relativamente aos reconhecimentos oculares e nasais. O reconhecimento dos olhos e do nariz é efectuado em cada face encontrada, da mesma forma que o reconhecimento de faces. Contudo, requer outro tipo de configuração para melhorar a obtenção de resultados. Tal como no reconhecimento facial, ambos os olhos e o nariz recebem coordenadas rectangulares no caso de reconhecimento, sendo estas utilizadas para a criação de círculos e linhas.

b) Armazenamento de Dados/Informações

O armazenamento dos dados nas listas é efectuado assim que todas as detecções faciais, oculares e nasais estejam concluídas. Cada face identificada é analisada através das coordenadas x e y do canto superior esquerdo do rectângulo que representa a detecção. A análise efectuada pelas aplicações a este ponto permite determinar se a face encontrada na *frame* actual já foi reconhecida numa *frame* anterior. Esta análise é feita através do factor proximidade. Assim, as coordenadas da detecção actual são comparadas com as coordenadas das detecções anteriores, efectuando-se a respectiva subtracção. Esta subtracção é efectuada apenas nas últimas coordenadas de cada face existente na lista, determinando-se se o valor da diferença é ou não muito elevado. Caso o valor seja superior a 20, nas coordenadas x ou y, relativamente a todas a faces contidas na lista, isto significa que a detecção actual está afastada das detecções encontradas nas *frames* anteriores, fazendo com que as aplicações criem um novo nó no final da lista de faces (detectou uma nova face) e criem também uma lista de coordenadas associadas à nova face. Se o valor for inferior a 20, trata-se de uma face já detectada anteriormente, pelo que as coordenadas serão guardadas nessa face, sendo criado somente um novo nó na lista de coordenadas e fazendo-se o incremento de um valor ao número de pontos pertencentes à face. Ainda no sentido de optimizar o armazenamento de faces para que as coordenadas de uma face não sejam colocadas numa face diferente, criou-se uma variável do tipo contador que permite controlar se uma pessoa deixou de aparecer nas *frames* seguintes ou não. Cada face, após ser inserida na lista de faces, herda um contador designado por "fim". Este contador é incrementado se não existirem novas coordenadas para adicionar à face após a análise de uma *frame*. Se o valor do "fim" for igual ou superior a10 *frames*, a face associada deixa de obter novas coordenadas, já que é muito provável que essa pessoa tenha abandonado o espaço em causa.

A informação que é armazenada relativamente aos olhos e ao nariz não requer uma procura pela face correcta, já que, dado que a respectiva detecção apenas é feita após a face ser detectada, as coordenadas dos olhos e do nariz ficam com a sua informação associada a essa face. Apesar desta associação, a informação recolhida acerca dos olhos só é guardada após ser analisada. Isto deve-se à necessidade de determinar a que olho pertence a detecção feita pelos classificadores (esquerdo ou direito). Apesar da utilização de dois ficheiros com classificadores, um destinado à detecção do olho esquerdo e o outro ao olho direito, pode acontecer que por vezes detectem ambos os olhos. Assim, a determinação do olho esquerdo ou direito é feita de forma simples através da obtenção da linha central da janela de pesquisa, sendo classificadas como olho esquerdo todas as detecções obtidas antes do centro e como olho direito as obtidas após o centro. De forma a optimizar estas detecções foi necessário estabelecer prioridades de armazenamento dos dados, pois, nos casos em que um ficheiro de classificadores obtém a detecção de ambos olhos foi necessário estabelecer quais é que seriam guardados. Portanto, como o reconhecimento do olho esquerdo se inicia em primeiro lugar, as detecções assim obtidas serão armazenadas na lista de coordenadas correspondentes à face. Contudo, as coordenadas que foram recolhidas relativamente ao olho direito podem vir a ser substituídas pelas detecções dos classificadores do olho direito. De facto, dado que os classificadores estão dedicados à obtenção do olho direito, estes têm necessariamente maior prioridade do que a detecção anterior. Desta forma, em casos para os quais os classificadores não identificam o olho a que se destinam, estes podem ser auxiliados pelos outros classificadores. Para além disso, para que os dados obtidos possam corresponder realmente ao olho de uma pessoa, foi feita uma filtragem pela localização dos olhos. Assim, todos os valores obtidos que não se encontram dentro da área estabelecida que contém os olhos serão eliminados. Concluindo o armazenamento dos dados, os valores obtidos para a localização do nariz são guardados após a obtenção das coordenadas fornecidas pelos classificadores, pois os dados referentes ao nariz não necessitam de análise adicional.

c) Determinação das Orientações

A determinação das orientações faciais em imagens (2D) necessita de pelo menos três pontos de referência para a correcta determinação da orientação de uma face. Para este efeito, optou-se pela escolha dos pontos recebidos nas detecções de ambos os olhos e do nariz. A obtenção das orientações é feita através do cálculo da distância das coordenadas no eixo x de ambos os olhos relativamente ao valor x do nariz. A figura 4 ilustra a detecção dos dois tipos de orientação da face: quando uma pessoa possui uma dada orientação, a identificação do nariz surge cada vez mais próxima do olho correspondente a essa orientação. Isto significa que nas orientações à esquerda a distância entre o nariz e o olho esquerdo é menor que a distância entre o nariz e o olho direito, como se mostra no primeiro caso da figura 4 (imagem da esquerda). Para o segundo caso ilustrado na figura 4 (imagem da direita), verifica-se precisamente o inverso - um valor inferior para a diferença entre o olho direito e o nariz, determinando que a orientação da face é para a direita.

Figura 4. Orientação à Esquerda/Orientação à Direita

2) Visualização das Frames e Dados

A fase de visualização da *frame* permite que o utilizador possa, caso pretenda, retirar notas que lhe possam ser úteis durante a execução. Para cada pessoa detectada, ficam visíveis as áreas que representam a face, os olhos, o nariz e apresentam ainda o ID correspondente a cada face detectada. Para além disso, é possível visualizar as orientações que foram tomadas pelas pessoas, sendo cada orientação associada à pessoa a que pertence através do respectivo ID de face. Na figura 5 é possível verificar a apresentação destes dados que são disponibilizados durante a visualização de uma *frame*.

Figura 5. Visualização de uma Frame e Dados

3) Eliminação de Faces

A fase da eliminação de faces é importante para eliminar as falsas detecções que forem obtidas e armazenadas na lista durante a análise da *frame*. De facto, após se ter realizado uma análise exaustiva em vários vídeos guardados, verificou-se que as falsas detecções surgiram somente em duas ou três *frames* de forma consecutiva. Tendo em conta esta observação, determinou-se que as faces armazenadas na lista seriam eliminadas ao fim de cinco *frames* após a sua inserção na lista de faces caso não obtenham pelo menos quatro novos nós de coordenadas. Para além disso, este procedimento ocorre não só para as falsas detecções como também para todas as faces que se encontram com menos de quatro pontos em cada grupo de cinco *frames* analisadas.

4) Menu de Resultados

Finalmente, o menu de resultados que surge no final da visualização do vídeo permite ao utilizador a consulta de todas as detecções feitas durante o decurso do vídeo, ou durante a recolha de vídeo em tempo real. O utilizador recebe informação sobre o número total de faces classificadas como diferentes que foram detectadas podendo visualizar o trajecto de uma face ou verificar o número de orientações de face que foram reconhecidas. A informação completa sobre as faces fica disponível assim que o utilizador escolha a face pretendida. Tal informação diz respeito ao número de *frames* que passaram após ter sido feita a primeira detecção, o número de detecções obtidas para a respectiva face e, finalmente, todas as orientações que foi possível identificar ao longo de todo o percurso, juntamente com o intervalo de tempo tomado pela face em cada orientação. É ainda possível verificar qual foi o trajecto tomado pela pessoa na área abrangida pela câmara, assinalado por uma linha vermelha juntamente com os pontos onde ocorreu a primeira identificação de uma orientação e as mudanças de orientação que foram tomadas ao longo do tempo.

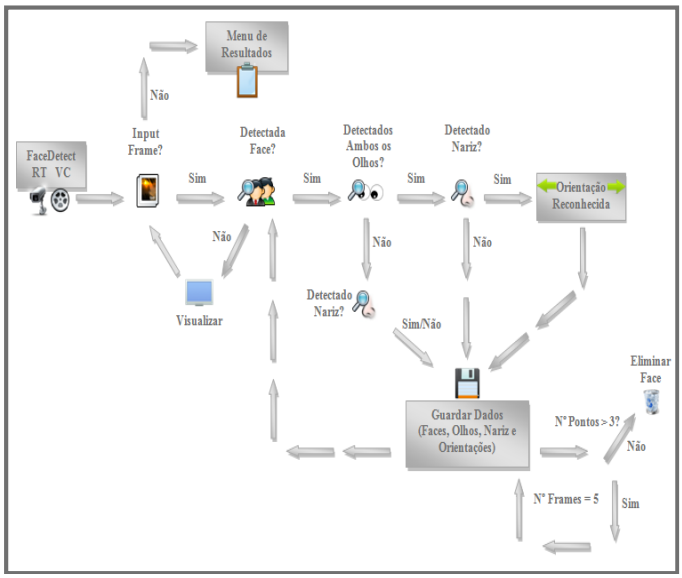

Figura 6. Diagrama de Fluxo

5) Diagrama de Fluxo

Na figura 6representa-seo diagrama de fluxo de ambas as aplicações, resumindo o respectivo processamento. As aplicações iniciam com o envio de uma frame para análise de forma a determinar se existe alguma face presente. Em caso de sucesso, para cada face encontrada realiza-se uma pesquisa a fim de detectar a região dos olhos, seguida da pesquisa pela área do nariz. Se adquirirem os três pontos de referência necessários (ambos os olhos e nariz), então inicia-se a determinação da orientação da cabeça, ou face, dessa pessoa, terminando com o armazenamento de todos os dados adquiridos sobre a face, os olhos, o nariz e a respectiva orientação. Segue-se a função de eliminação que verifica se a

face adicionada cumpre os requisitos mínimos. O ciclo repete-se então para a face seguinte até não existirem mais faces. Mostra-se então ao utilizador as detecções adquiridas na frame actual e adquire-se mais uma frame para análise. No fim, a aplicação termina e apresenta o menu de resultados.

IV. AVALIAÇÃO

Para avaliara aplicação *FaceDetect*, analisou-se e avaliou-se o desempenho dos classificadores para as respectivas detecções, o armazenamento dos dados e as orientações reconhecidas correctamente pelas aplicações. Sendo a análise do desempenho dos classificadores *Haar* um processo complexo devido ao facto de ser necessário avaliar, entre outros parâmetros, (i) as várias etapas dos classificadores, e (ii) as várias dimensões dos objectos a serem detectados, optou-se por considerar resultados mencionados na literatura que permitem aferir o grau de desempenho dos classificadores que foram utilizados no desenvolvimento das aplicações. Relativamente à avaliação do armazenamento das faces e das orientações que foram correctamente obtidas, optou-se pela utilização de um ficheiro de vídeo guardado em disco cujas características principais se identificam na tabela II.O vídeo de teste apresenta um ambiente não controlado no qual é possível identificar as faces de pessoas, bem como as respectivas orientações de face durante a passagem num espaço físico.

Tabela II. CARACTERÍSTICAS PRINCIPAIS DO VÍDEO DE TESTE

Espaço ocupado	**Resolução**	**Taxa de *frames***	**Duração do vídeo**	**Número de *frames***
2.26 MB	768×576	20 FPS	12 s	244 *frames*

De acordo com Castrillón-Santanaetal.[10], vários autores optaram por criar e tornar os seus próprios classificadores públicos desde a publicação do mais recente trabalho de Viola & Jones[11]. De forma a analisar o desempenho dos vários classificadores, os autores utilizaram o conjunto de dados da CMU [12] divididos em quatro categorias diferentes: *test-low*, *test*, *rotated* e *newtest*, sendo cada categoria uma combinação dos dados fornecidos em[13]e[14]. O conjunto é composto por várias imagens *grayscale* que contêm faces diferentes com diversas orientações. Para testar o desempenho dos classificadores calculou-se a taxa de detecções obtidas por número de falsos positivos, tendo obtido resultados bastante promissores[10], quer ao nível de detecções oculares quer às nasais. Para determinar o grau de acertos das detecções nas aplicações desenvolvidas optou-se por registar todas as falsas detecções e todas as eliminações que sucederam ao longo do processamento do vídeo de teste, através da análise visual do vídeo e dos dados armazenados nas listas, tendo-se obtido os resultados indicados na tabela III. Ao longo de todo o processamento identificaram-se e guardaram-se treze faces, sendo cada uma delas considerada como pertencente a uma pessoa diferente. Contudo, no vídeo de teste estavam apenas presentes sete pessoas. Isto sucede nos casos em que, apesar de já ter sido detectada e armazenada uma pessoa nas *frames* anteriores, a mesma pessoa volta a ser armazenada como uma nova face. Esta situação ocorre sempre que uma pessoa deixa de ser detectada e só volta a ser detectada mais tarde, mas também pode ocorrer devido ao posicionamento da câmara, nos casos em que a câmara de videovigilância se encontra muito perto do nível do solo ou numa posição que permite uma grande aproximação de uma pessoa à câmara. Nestes casos, o armazenamento da mesma pessoa como uma nova face ocorre mais frequentemente, pois quanto mais próxima da câmara estiver a pessoa, maior será a diferença entre as coordenadas do ponto que foi armazenado dessa face nas *frames* seguintes.

Tabela III. RESULTADOS DO ARMAZENAMENTO DOS DADOS

Falsas Detecções	**Falsas Detecções Eliminadas**	**Faces Eliminadas com Menos de 3 Pontos**
11	11	19

Na obtenção dos resultados relativamente às orientações foi necessário identificar todos os casos para os quais foi possível obter os três pontos de referência (dois olhos e nariz) e determinar se as orientações obtidas pelas aplicações correspondem de facto às orientações tomadas pelas pessoas no vídeo de teste. Os resultados da tabela IV demonstraram que, tendo detectado os três pontos de referência, as aplicações conseguem determinar as orientações de uma pessoa com uma taxa de sucesso de 77.9%. Esta taxa poderá vir a aumentar se as detecções efectuadas pelos classificadores para os olhos e para o nariz também aumentarem, já que, nos casos onde ocorreram orientações incorrectas, verificou-se que sucederam devido a uma falha na detecção do olho esquerdo ou do direito. A falha diz respeito a situações em que um dos olhos fica fora da zona correcta do olho, fazendo com que o respectivo olho tenha uma maior proximidade com o nariz do que o olho oposto, dando assim uma orientação incorrecta para essa pessoa.

Tabela IV. RESULTADOS DAS ORIENTAÇÕES

Orientações Reconhecidas	**Orientações Correctas**	**Orientações Incorrectas**
95	74	21

Para averiguar o tempo total do processamento da aplicação *FaceDetectVC*, submeteu-se, em primeiro lugar, o vídeo de teste a vários *frame rates*. Para a segunda avaliação manteve-se o *frame rate* mas alterou-se a resolução do vídeo. Os resultados obtidos ilustram-se na tabela V. Como seria de esperar, verificou-se que à medida que aumenta o *frame rate*, aumenta igualmente o tempo que demora a análise do vídeo completo. Em contrapartida, obtém-se um maior número de detecções. O valor do tempo de execução indicado diz respeito à duração da análise de um número de *frames* igual ao valor do *frame rate*. Para um vídeo de 15 fps, a aplicação demorou cerca de 10 s.

Tabela V. RESULTADOS DO VÍDEO DE TESTE COM DIFERENTES FPS

FPS (*frame rate*)	**15 FPS**	**20 FPS**	**23 FPS**	**25 FPS**	**30 FPS**
Tempo de Execução	10 Seg.	12 Seg.	14 Seg.	16 Seg.	19 Seg.
Faces Encontradas	12 Faces	13 Faces	15 Faces	17 Faces	18 Faces
Nº Total de Detecções Guardadas	194	274	359	367	450

Para determinar se a resolução de um vídeo iria afectar as detecções efectuou-se ainda uma análise com várias resoluções diferentes. O vídeo contém 20 fps e a sua resolução original é de 768×576. Os resultados obtidos para diferentes resoluções

mostram-se na tabela VI, confirmando que a resolução de um vídeo afecta a velocidade do processamento das *frames* e também as respectivas detecções. Neste caso, verificou-se que resolução de 720×480 foi a que menos afectou o desempenho das detecções e obteve um menor tempo de processamento em relação ao vídeo original. Ainda é possível verificar que as detecções são mais afectadas se os valores de resolução da imagem, na vertical e na horizontal, não forem proporcionais. Por isso, estes valores devem manter, quanto possível, a imagem próxima da realidade. Por exemplo, no caso da resolução 720×576 a detecção foi inferior às detecções com a resolução 720×480 devido à ligeira distorção da imagem.

Tabela VI. RESULTADOS DO VÍDEO DE TESTE COM DIFERENTES RESOLUÇÕES

Resolução	**352×240**	**352×288**	**720×480**	**720×576**	**768×576 original**
Tempo de Execução	2.6 Seg.	3.1 Seg.	10 Seg.	11 Seg.	12 Seg.
Faces Encontradas	5 Faces	5 Faces	12 Faces	10 Faces	13 Faces
Nº Total de Detecções Guardadas	73	99	271	235	274

Finalmente, a aplicação *FaceDetectRT* foi testada com uma simples *webcam* com aproximadamente 1.3 megapixéis, uma resolução de 352×240 e *frame rate* de 15fps. Neste caso mediu-se apenas o atraso que a aplicação evidencia durante as detecções feitas em tempo real. Esses resultados foram bastante positivos tendo-se obtido um atraso entre 0.9 e 1.4 segundos.

V. CONCLUSÃO

O objectivo inicial de desenvolver aplicações para efectuar a detecção e o seguimento automático de pessoas em ambientes não controlados foi atingido. A recolha e análise dos resultados dos testes efectuados possibilitaram a recolha de informação importante, que contribui para determinar a precisão e o desempenho das aplicações na obtenção dos dados sobre as pessoas no ambiente abrangido pelas câmaras de vídeo. Em termos da precisão das aplicações, efectuamos um balanço igualmente positivo pois, apesar do tempo de processamento ser um pouco elevado, as aplicações detectaram deslocações correctas das pessoas nos ambientes não controlados, o que significa que essas pessoas foram detectadas na orientação correcta, evitando as falsas detecções que ocorreram ao longo do processamento, obtendo-se uma taxa de orientações correctas de aproximadamente 78%.Uma das limitações das aplicações descritas neste trabalho tem a ver com o facto de, nos casos para os quais não é possível obter os três pontos de referência (olhos e nariz) de uma face, não ser possível determinar a orientação de uma dada pessoa. Destaca-se igualmente a dependência que as orientações possuem dos resultados obtidos pelos classificadores na obtenção das posições de ambos os olhos e do nariz: caso as posições estejam incorrectas, provocam a incorrecção da orientação.

REFERÊNCIAS

[1] Y. Benezeth, B. Emile, H. Laurent, and C. Rosenberger, "Vision-Based System for Human Detection and Tracking in Indoor Environment," International Journal of Social Robotics, vol. 2, Dec. 2009, pp. 41-52.

[2] K. Kollreider, H. Fronthaler, M.I. Faraj, and J. Bigun, "Real-Time Face Detection and Motion Analysis With Application in 'Liveness' Assessment," IEEE Transactions on Information Forensics and Security, vol. 2, Sep. 2007, pp. 548-558.

[3] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2001.

[4] C.P. Papageorgiou, M. Oren, and T. Poggio, "A general framework for object detection," International Conference on Computer Vision, 1998, pp. 555-562.

[5] Y. Freund, "A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting," Journal of Computer and System Sciences, vol. 55, Aug. 1997, pp. 119-139.

[6] R. Lienhart and J. Maydt, "An extended set of Haar-like features for rapid object detection," Proceedings. International Conference on Image Processing, 2002, p. I-900-I-903.

[7] Intel, "Open Computer Vision Library."

[8] G. Bradski and A. Kaehler, Learning OpenCV: Computer vision with the OpenCV library, O'Reilly Media, 2008.

[9] A. Harvey, "OpenCV Face Detection: Visualized," 2010.

[10] M. Castrillón-Santana, O. Déniz-Suárez, L. Antón-Canalís, and J. Lorenzo-Navarro, "Face and facial feature detection evaluation," Third International Conference on Computer Vision Theory and Applications, VISAPP08, 2008.

[11] P. Viola and M. Jones, "Robust Real-Time Face Detection," International Journal of Computer Vision, vol. 57, 2004, pp. 137-154.

[12] H. Schneiderman and T. Kanade, "A statistical method for 3D object detection applied to faces and cars," Proceedings IEEE Conference on Computer Vision and Pattern Recognition CVPR 2000 Cat NoPR00662, vol. 1, 2000, pp. 746-751.

[13] K.-K. Sung and T. Poggio, "Example-based learning for view-based human face detection," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 20, 1998, pp. 39-51.

[14] H.A. Rowley, S. Baluja, and T. Kanade, "Neural network-based face detection," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 20, 1998, pp. 23-38.