Detecção de Faces e Rastreamento do Olhar em Imagens

Flavio Luiz Coutinho

MAC 5701

Tópicos em Ciência da Computação

Relatório de Estudos

Orientador: Carlos Hitoshi Morimoto

1 Introdução

Com a evolução das novas tecnologias de informação e mídia, novas abordagens para a interação homem-computador estão sendo desenvolvidas muitas vezes elas não contam mais com os dispositivos tradicionais de entrada e saída (como teclado, mouse e monitor). A evolução da tecnologia e a redução de custos para a aquisição e processamento de imagens digitais tornou possível a aplicação da visão computacional como dispositivos alternativos de entrada e saída nos sistemas existentes.

O processamento de faces é um tópico que vem sendo bastante estudado nos últimos anos. O objetivo é que informações sobre a identidade, estado e intenções de um usuário possam ser extraídas a partir de imagens e que os sistemas respondam a essas informações de forma adequada. O primeiro passo para qualquer sistema de reconhecimento de faces é realizar a detecção das regiões em uma imagem nas quais existam faces presentes. Contudo essa tarefa é um tanto difícil devido as inúmeras variações que os dados de entrada (ou seja, as imagens) podem apresentar, aumentando de forma significativa a complexidade do problema. Obviamente pode-se fazer certas restrições para atacar o problema em um universo reduzido, o que, apesar da perda da generalidade da solução, pode resultar em resultados mais precisos.

Como ponto de partida, vamos definir de forma mais objetiva o termo *detecção de faces*. Dada uma imagem, a meta da detecção de faces é determinar se há ou não alguma face na imagem e em caso positivo delimitar as regiões de ocorrências das faces. Existem ainda uma série de outros problemas que estão relacionado a detecção de faces. A *localização de faces*, por exemplo, consiste em determinar a posição de uma face em uma imagem dado que realmente existe uma face na imagem. Isto é, é uma simplificação do nosso problema de detecção de faces. Outro problema relacionado é o *reconhecimento de faces* ou *identificação de faces*. Neste problema dada uma imagem de entrada ela é comparada a um banco de dados e é relatado se existe algum casamento com alguma imagem do banco. Semelhante a este último temos também a *autenticação de faces* que tem como propósito verificar a identidade de um indivíduo em uma imagem. Já o *rastreamento de faces* adiciona a dimensão do tempo ao nosso problema de detecção e estima de forma contínua e em tempo real a localização de uma face em uma seqüência de vídeo. Finalmente o *reconhecimento de expressões faciais* envolve a identificação de estados emotivos como felicidade, tristeza, etc.

Neste trabalho foram realizados estudos de algumas técnicas de visão computacional empregadas para resolver o problema da detecção de faces e também um pouco de rastreamento do olhar de faces em imagens. O estudo dos primeiro problema fornece uma base de técnicas e ferramentas que são utilizadas a fim de se resolver o segundo problema. A seguir veremos com maior detalhes cada um desses assuntos.

2 Detecção e Rastreamento de Faces em Imagens

Seja qual for o problema, dentre os acima citados, a detecção de faces constitui o primeiro passo a ser tomado para a implementação de qualquer solução automatizada para esses problemas. Uma idéia inicial, baseada no conhecimento que cada um de nós possui sobre faces, seria estabelecer algumas regras simples para descrever características de uma face e a relação existente entre essas características. Assim, uma face vista de frente é composta de dois olhos, um nariz e uma boca. Além disso as essas características devem estar dispostas de forma que os olhos estejam em posições simétricas, o nariz e a boca estejam no eixo perpendicular ao eixo passando entre os dois olhos e o nariz esteja localizado entre a boca e os olhos. Ou seja, primeiro faz-se um levantamento de possíveis características e depois faces candidatas são identificadas baseada nas regras definidas, acompanhado geralmente de um processo de verificação para evitar detecções falsas.

Um problema com essa idéia inicial é que é difícil expressar esse conhecimento que cada um de nós possui em termos de regras bem definidas. Se as regras forem muito exigentes podemos deixar de detectar faces, se as regras forem muito genéricas podemos detectar outras coisas que não sejam faces. Além disso existem uma série de fatores que afetam a complexidade deste problema (pois variam demais a aparência que uma face pode ter) e constituem grandes desafios. São eles:

- pose: orientação da face em relação a câmera, pode ser frontal ou de perfil por exemplo. Algumas características faciais (como os olhos ou nariz) podem ficar parcialmente ou totalmente ocultos em diferentes poses.
- presença de características particulares: barba, bigode e óculos são exemplos de características particulares que podem aparecer de forma bastante variada na cor, formato e tamanho.
- expressão facial: a aparência das faces são diretamente afetadas pelas expressões faciais de uma pessoa.
- regiões ocultas: regiões da face podem estar ocultas por objetos ou mesmo outras faces (no caso de um grupo de pessoas).
- orientação da Imagem: rotação das face ao longo do eixo óptico da câmera.
- condições da imagem: fatores como iluminação e características do dispositivo de aquisição (lentes, resposta de sensores).

Para lidas com essas complicações muitos estudos e pesquisas foram realizadas e escolhi algumas abordagens propostas para estudar mais detalhadamente. A primeira deles [4] baseia-se na detecção de faces baseadas em características. A idéia é que existam características de faces invariantes a diversas condições de iluminação e variação de pose e que essas características uma vez detectadas possam ser utilizadas para inferir a presença de uma face. A segunda [3], [6] e [7] baseia-se na utilização de um modelo de cores fazer a detecção e também o rastreamento de faces em imagens. Já a terceira abordagem estudada [2] propõe um método de detecção de faces que usa iluminação ativa e explora uma propriedade reflexiva dos olhos.

2.1 Abordagem Baseada em Características Faciais

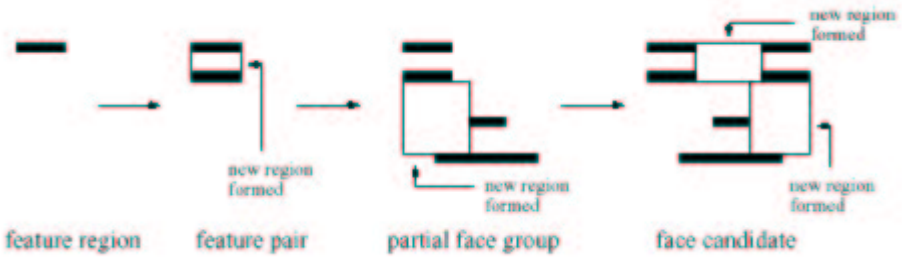
Neste trabalho, a face é modelada como um plano contendo seis características faciais orientadas (sombrancelhas, olhos, nariz e boca). Além disso, as regiões correspondentes às bochechas precisam não devem apresentar outros pontos candidatos a serem consideredos características faciais. Este modelo de face tem as vantagens de usar pontos interiores a face (pois caso fossem usados pontos na região da borda da face, estes poderiam ser facilmente influenciados pelos cenário de fundos da imagem) e de cobrir uma boa área da face (comparado outro modelo desenvolvidos que utilizam como características os olhos, narinas e ponta do nariz). Esse modelo de face é também decomposto em componentes cada uma contendo quatro características. Essas componentes representam ocorrências comuns de faces sob diferentes poses. Essas grupos são chamados *Partial Face Groups*. Esses grupos são ainda divididos mais uma vez em grupos menores contendo duas características (pares horizontais e pares verticais – *Hpairs* e *Vpairs*). Na figura abaixo podemos ver o modelo de faces e a divisão em Partial Face Groups e ainda Hpairs e Vpairs:

Para que a detecção das características faciais seja robusta, é necessário que sejam usadas características da imagem que sejam invariantes a escala e iluminação. Foi observado que em baixas resoluções todas as seis características faciais apresentadas acima aparecem como linhas alongadas sobre uma região clara da face. E como bordas são invariantes a iluminação dentro de uma certa escala, as seis características faciais são modeladas como pares de arestas orientadas. A imagem de entrada é suavizada antes de ocorrer a detecção de características a fim de que características de alta resolução tomem a forma das características de baixa resolução. Veja na figura abaixo os modelos utilizados para as seis características faciais:

Claramente, com um modelo que busca características de baixo nível (pares de arestas orientadas), haverá uma quantidade grande de características detectadas que não pertencem a uma face. Para lidar com isso, foi proposto um modelo para agrupamento por percepção que junta as características encontradas utilizando informações geométricas e relativas a tonalidade de cinza dos pontos.

Em linhas gerais o processo de detecção de faces proposto funciona da seguinte maneira: uma lista de pontos de interesse são extraído da imagem de entrada (a partir da caracterização das características faciais apresentadas acima). Arestas em torno de cada ponto de interesse são examinadas e agrupadas baseado na proximidade, similaridade na orientação e força. Assumindo faces na posição vertical, procura-se por arestas aproximadamente horizontais acima e abaixo do ponto de interesse (ou se a orientação da face é desconhecida, pode-se procurar por arestas aproximadamente paralelas e com polaridade opostas em relação ao ponto de interesse). Achando tais arestas, o ponto é marcado como uma possível característica facial, e a região da característica é delimitada. São feitas medidas das características da região e guardadas em um vetor de características **x**. A partir das imagens de treino são definidos os valores médios e matrizes de covariância para cada categoria de características faciais (sombrancelhas, olhos, nariz e boca), e o vetor **x** validado conforme esses dados de treino. Aqui acaba a fase de seleção de características do processo e inicia-se a fase de agrupamento das características encontradas.

Depois de obter uma lista de pontos validos e suas regiões associadas, essas regiões características são agrupadas usando o modelo face já discutido. Características simples são agrupadas em Hpairs e Vpairs. Esses pares são por sua vez agrupados em Partial Face Groups e finalmente esses grupos são agrupados em candidatos a faces. Além disso, para cada fase do agrupamento, vetores de características são extraídos e validados conforme medidas obtidas das imagens de treino. A figura abaixo ilustra o processo de agrupamento de características em grupos maiores até chegarmos a uma face candidata:



2.2 Abordagem Baseada em Modelo de Cores

Outra abordagem utilizada para a detecção de faces em imagens [3] é realizar a localização através da procura por regiões na imagem que contenham pixels na tonalidade da cor da pele. Assim como olhos, nariz e boca são considerados características no processo de detecção de faces, a cor também pode ser considerada uma característica de face. Além disso, usar a cor da pele para detectar faces possui as seguintes vantagens: processar cores é mais rápido do que processar outras características faciais; não há o problema de a característica poder estar oculta (ou estar visível e ficar oculta no caso de rastreamento de faces em seqüências de imagens); e finalmente sob certas condições de iluminação a cor é invariante a rotações. Esta ultima propriedade é particularmente interessante quando se está fazendo rastreamento de faces e se quer estimar o movimento realizado pela face pois a movimentação pode ser modelada usando um modelo de translação (da forma $X' = X + b$, $b \in R^2$) que é mais simples do que modelos baseados em transformações afins e transformações projetivas.

Entretanto o uso da cor como característica apresenta também o seus problemas. Um desses problemas está no fato de que a representação de cor obtida por uma câmera é influenciada por muitos fatores como luz ambiente por exemplo. Outro problema é que câmeras diferentes produzem valores de cores significativamente diferentes mesmo para uma mesma cena em condições iguais. Por último, a cor de pele humana difere de pessoa para pessoa. Assim, se quisermos usar a cor como uma característica na detecção e rastreamento de faces, estes problemas precisam ser contornados de alguma forma. Muitas pesquisas foram feias a fim de contornar os problemas mostrados acima. O especial o modelo proposto em (referencia) é facilmente adaptável para pessoas diferentes e diferentes condições de iluminação.

O modelo desenvolvido é um modelo estatístico para cor de pele que trabalha no espaço cromático de cores. O Motivo para se trabalhar neste espaço de cores é que no espaço de cores RGB (adotado por boa parte das câmeras) cada valor de pixel, da forma (r, g, b), representa não apenas a informação de cor, mas também a informação de brilho do ponto representado na imagem. Como o brilho não é importante para caracterizar a cor da pele (sob condições normais de iluminação) pode-se representar a cor da pele no espaço cromático de cores (também conhecido como “cores puras” na ausência de brilho). O valor de um pixel (R,G,B) é definido da seguinte maneira no espaço de cores cromáticas:

- $r = R/(R + G + B)$
- $g = G/(R + G + B)$

Note que a cor azul é não é calculada pois pode ser obtida da seguinte forma: $b = 1 – r – g$. Dessa forma, para representar um pixel nesse espaço de cores precisamos apenas de dois valores (r, g), isto é, reduzimos a dimensão do espaço de cores inicial (RGB).

Pesquisas e estudos realizados mostraram a distribuição dos tons de cor de pele de uma face, no espaço de cores cromáticas, estão agrupadas em uma pequena área desse espaço de cores. Isto é, apenas algumas poucas cores dentre as existentes no espaço ocorrem em uma

face. Além disso, foi também verificado que distribuições de tons de pele de pessoas diferentes e de raças diferentes também estão agrupadas no espaço cromático de cores. Isso porque apesar da grande variedade de tons de pele entre diferentes pessoas, as variações estão mais ligadas ao brilho do que a cor em si. Variações de iluminação podem afetar um pouco a distribuição de cores no espaço trabalhado mas a forma da distribuição se mantêm regular. Mais ainda, esse distribuição é similar a uma distribuição Gaussiana. Dessa forma, com todos esses resultados, uma distribuição de cores de pele pode ser representada por um modelo Gaussiano $N(m, \sum^2)$ onde m é a média e $\sum$ é a matriz de covariância dos tons de pele obtidos de uma amostra.

Para fazer a detecção da faces utilizando esse modelo, cada pixel da imagem de entrada é convertido para o espaço cromático e depois comparado com a distribuição de cores obtida pelo modelo, classificando assim o pixel como tendo cor de pele ou não. Contudo é possível que outras regiões de uma imagem contenham pixels com tonalidades de pele, como cenários de fundo ou mesmo mãos de pessoas. É impossível detectar faces simplesmente usando os resultados da comparação de cor. Na figura abaixo podemos ver uma exemplo aonde além das faces outras regiões foram segmentadas como tendo cor de pele:

Assim, informação adicional torna-se necessária para fazer uma correta detecção. Quando trabalhando com seqüências de imagens pode-se utilizar informações de movimento para detectar se uma região caracterizada com cor de pele é realmente uma face ou não. Além disso informações geométricas como tamanho e forma também podem ser utilizadas. Por exemplo, uma mão, apesar de ser da cor da pele, não possui nem o tamanho e nem o formato de uma face.

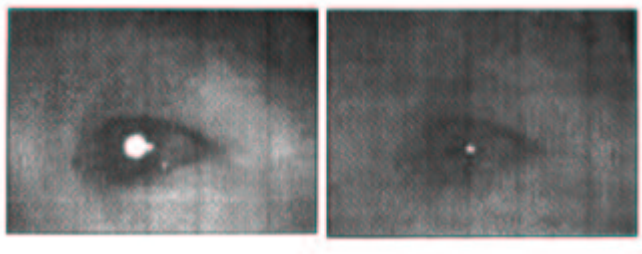
No trabalho proposto em [7], que é um sistema para rastreamento de olhar, este modelo de cores é utilizado para fazer a busca pela face. Como nesse sistema o rastreamento é feito para apenas uma face, a maior região conexa composta por pixels com cor de pele é

delimitada como sendo a região da face. Após a detecção dessa região, outras características faciais como olhos, narinas e boca são determinadas. A determinação dos olhos por exemplo é feita pela procura de duas regiões escuras na parte superior da região delimitada pela face. A procura pelo narinas é feita da mesma forma do que a dos olhos, mas a região de procura é a região compreendida entre os olhos e a boca. A procura pela boca por sua vez é um pouco mais elaborada mas em linhas gerais procura por um segmento escuro e aproximadamente horizontal. Neste outro trabalho desenvolvido em [6] que detecta múltiplas faces em imagens, também é utilizado um modelo de cor de pele semelhante ao visto acima para detectar as regiões candidatas a faces. Essas regiões são depois agrupadas até formarem regiões aproximadamente elípticas e são ainda validadas posteriormente. Este trabalho se mostrou bem sucedido na detecção independente do tamanho, orientação e pose das faces.

2.3 Abordagem Baseada em "Active Ilumination"

Uma terceira abordagem, diferente das duas anteriores já vistas, é baseada em uma técnica robusta para detecção de pupilas que explora uma propriedade reflexiva dos olhos para facilitar a detecção. O sistema de detecção de faces proposto neste trabalho é composto por uma câmera preto e branco e duas fontes de luz. Cada fonte de luz é formada por grupos de leds infravermelhos. O primeiro conjunto de leds é posicionado bem próximos ao eixo óptico da câmera. Já o segundo grupo de leds é dividido em dois subgrupos e colocados simetricamente em cada lado do eixo óptico da câmera.

Devido a propriedade reflexiva do olho, que diz que a luz infravermelha é refletida exatamente ao longo da direção pela qual ela incide no olho, quando o sistema acima descrito utiliza sua primeira fonte de iluminação as imagens obtidas pela câmera apresentam pupilas brilhantes. Quando a fonte de iluminação é utilizada as imagens obtidas apresentam as pupilas mais escuras mas brilho similar para todo o resto da imagem. Na figura abaixo podemos ver os dois tipos de imagens obtidas com o uso de cada fonte de luz:



Essa propriedade reflexiva dos nossos olhos são difíceis de serem encontradas em outros objetos de forma que as pupilas geralmente são os únicos objetos que aparecem com um alto contraste comparando-se as imagens obtidas com cada uma das condições de iluminação. Assim o simples processo de aplicar um threshold à diferença entre uma

imagem com a pupila brilhante e uma imagem com a pupila escura é suficiente para detectar as pupilas de uma face presente na imagem. Óculos e lentes de contato não alteram essa característica reflexiva dos olhos e portanto não inibem a detecção utilizando esse método. Reflexos de óculos podem atrapalhar o método mas na pratica essas condições não duram por mais de algumas frações de segundo o que chega a atrapalhar o rastreamento das pupilas. Na figura abaixo podemos observar o resultado obtido pela aplicação de um threshold à diferença entra as imagens mostradas na figura anterior:

Detalhando um pouco mais o funcionamento do sistema, foi desenvolvido um mecanismo para sincronizar o acionamento de cada uma duas fontes de luz à varredura das linhas pares ou impares pela câmera (uma vez que o sinal de vídeo é composto por frames intercalados contendo linhas pares e linhas impares captadas alternadamente). Assim, para um frame obtido pela câmera (Ft), temos um frame correspondente linhas pares (Et) e um frame correspondente as linhas impares (Ot). Et é a imagem em que as pupilas aparecem brilhantes enquanto Ot é a imagem na qual as pupilas aparecem escuras. Aplicando uma operação de threshold sobre Dt = (Et – Ot), fazendo em seguida uma rotulação das componentes conexas e finalmente verificando (através de critérios referentes a forma e tamanho) as componentes encontradas, são obtidos os candidatos a pupilas.

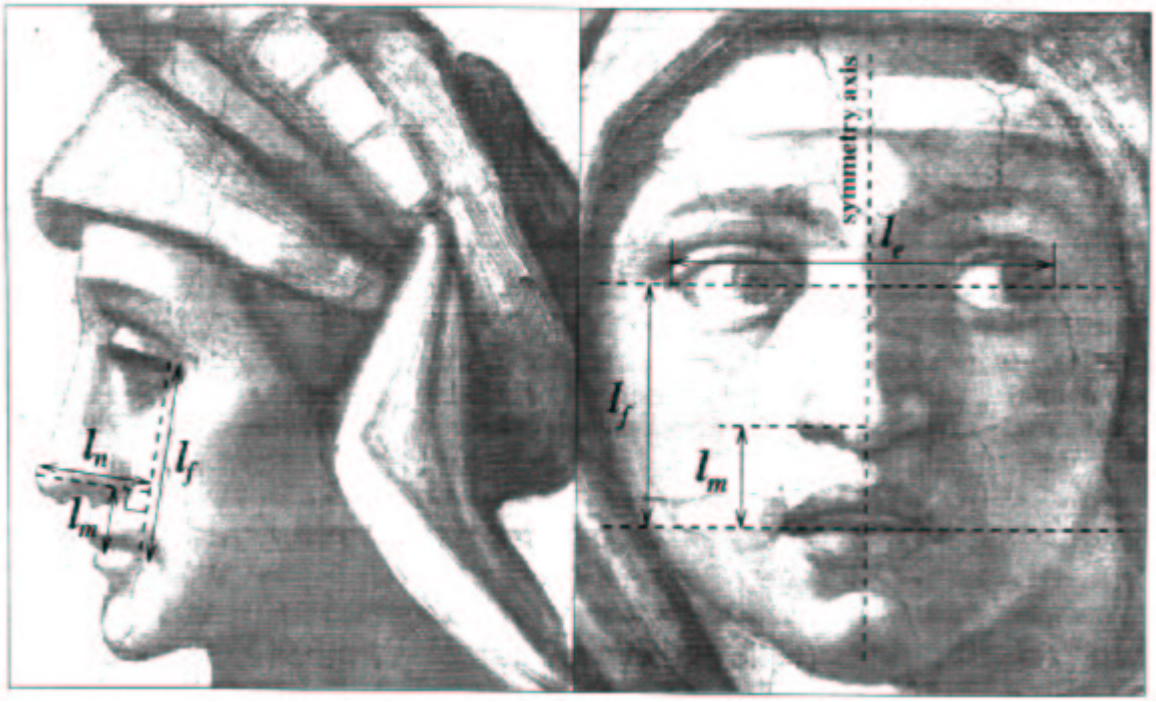
A próxima etapa é fazer a detecção de faces uma vez que os candidato a pupila foram encontrados. No caso mais simples de uma única face presente, a posição da face pode ser determinada assim que duas pupilas são detectadas. Mas é necessário técnicas mais elaboradas quando existem outras faces. Algumas heurísticas simples podem ser usadas no agrupamento das pupilas em faces. Essas heurísticas podem tanto ser baseadas em informações espaciais quanto em informações temporais. No caso de informações espaciais, podemos utilizar propriedades como tamanho, posição, cor da íris, tonalidade de pele envolvendo os olhos, além do fato de olhos de uma mesma pessoa tendem a estar dispostos em uma linha aproximadamente horizontal e possuir o mesmo tamanho. Já no caso de informações temporais podemos utilizar o piscar de olhos (que ocorrem ao mesmo tempo para uma pessoa) e a movimentação dos olhos (olhos de uma mesma pessoa movimentam-se da mesma maneira).

3 Rastreamento do Olhar de Faces em Imagens

Uma possível aplicação das técnicas de detecção de faces vistas até agora é o desenvolvimento de sistemas que rastreiam o olhar. A informação de para onde uma pessoa está olhando pode ser útil para uma serie de aplicativos interativos. Como exemplo podemos criar dispositivos apontadores alternativos e sistemas de realidade virtual. As técnicas vistas até agora para a detecção de faces são importantes pois para que possa-se determinar a direção do olhar de uma pessoa, é necessário antes de mais nada detectar e rastrear a face dessa pessoa.

Segundo Gee e Cipolla em [5], quando pensamos de forma intuitiva em como determinar a orientação de uma face, pensamos em pistas do tipo: a rotação da cabeça no sentido vertical pode ser estimada pela visibilidade do queixo ou o topo da cabeça; a rotação horizontal pode ser estimada pela visibilidade da orelha ou ainda a posição dos olhos em relação ao contorno da face. Apesar de intuitivamente corretas, essas pistas são baseadas em características faciais que variam muito de pessoa pra pessoa. Pessoas podem ter queiro duplo ou cabelos compridos que escondam as orelhas. Além disso essas pistas são definidas de forma meio vaga sendo assim difíceis de serem detectadas. Por isso, é necessário um conjunto de informações geométricas precisas e facilmente extraídas das imagens e que fornecem estimativas confiáveis. Devido essa razão, o método desenvolvido nesse trabalho usa apenas medidas referentes aos olhos nariz e boca. É necessário também possuir algum tipo de modelo para a geometria 3D das faces.

O modelo facial utilizado é baseado na razão de quatro medidas faciais: Lf, Ln, Le e Lm. Essas medidas se mostram apropriadas pois são características estáveis, isto é, não variam muito com expressões faciais. Podemos ver na figura abaixo essas medidas do modelo facial:



O problema da determinação da direção do olhar, reduz-se a estimar o vetor normal a face. Dois métodos para o calculo dessa normal são propostos. Mas seja qual for o método, todos os dados necessários para calcular o vetor normal a face são obtidos pelas medidas faciais do modelo, além de medidas extraídas da imagem sendo analisada com base nas posições dos olhos, nariz e boca.

Outro trabalho, desenvolvido por Q. Ji e R. Hu [8], propõe uma abordagem diferente para estimar a direção do olhar de uma face. Motivados pela dificuldade em detectar as características faciais necessárias para estimar a direção para onde a face aponta sob diferentes condições de iluminação e pose, a forma de uma face é modelada como uma elipse, o que garante as seguintes vantagens: primeiro uma face humana pode ser modelada de forma razoavelmente precisa por uma elipse. Segundo elipses são preservadas sob transformações projetivas. Terceiro, elipses representam a informação da face de uma forma mais robusta a ruído. E finalmente, é correspondência 2D/3D pode ser estabelecida de uma forma simples. Os resultados obtidos mostraram-se bastante satisfatórios, bem como os resultados obtidos pelo método visto em [referencia Cipolla].

4 Conclusão

Através deste trabalho foi possível estudar uma série de soluções para o problema de detecção de faces a também ver um pouco das soluções desenvolvidas para o problema mais especifico da determinação da direção olhar em imagens. Procurei selecionar algumas soluções bem distintas umas das outras para mostrar a variedade de soluções existentes para esses problemas. Existem ainda muitos outros estudos e trabalhos realizados a cerca deste assunto e cada uma das soluções podem apresentar resultados melhores em algum domínio especifico da aplicação, embora esforços sejam feitos para tornar os métodos os mais independentes possíveis dos domínios da aplicação. Observa-se também uma forte tendência de sistemas que utilizam soluções combinadas para resolver estes problemas estudados.

5 Referencias

[1] Ming-Hsuan Yang, David J. Kriegman and Nerandra Ahuja. Detecting Faces in Images:A Survey.

[2] C. H. Morimoto and M. Flickner. Real-Time Multiple Face Detetion Using Active Ilumination.

[3] J. Yang, A. Weibel . Tracking Human Faces in Real Time.

[4] Kin Choong Yow and Roberto Cipolla. Featura Based Human Face Detection.

[5] A. H. Gee and R. Cipolla. Determing the Gaze of Faces in Images.

[6] Ming-Hsuan Yang and Nerandra Ahuja. Detecting Human Faces in Color Images.

[7] R. Stiefelhagen, J. Yang and A. Waibel. A Model-Based Gaze Tracking System.

[8] Qiang Ji and Rong Hu. Face Pose Estimation from a Face Ellipse