6
Resultados

Nesse capítulo são apresentados os resultados individuais para cada etapa do trabalho proposto em dois blocos principais. O primeiro bloco exibe os resultados referentes aos modelos treinados e às comparações das duas formas de estado inicial. O segundo bloco mostra resultados referentes à aplicação desenvolvida para testar na prática o método de reconhecimento de emoções proposto. É importante realçar que os resultados dos dois blocos são separados, ou seja, os resultados de acurácia apresentados para o ASM e para os SVMs não sofrem influência dos resultados encontrados para aplicação desenvolvida e vice-versa. De maneira mais detalhada, inicialmente são exibidos resultados do método de Reconhecimento de Emoções, ou seja, a extração de atributos faciais com o ASM e de classificação de emoções com o SVM. Para os resultados do SVM, primeiramente é discutido o banco de dados utilizado e depois o que alcançamos especificamente com os modelos que seguiram cada uma das abordagens de estado inicial, forma média e neutra, seguido de uma comparação entre eles. Depois são exibidos detalhes de um modelo que utiliza a forma média em todo seu potencial, isto é, com a adição da classe neutra para classificação. Por fim, o desempenho e considerações sobre o programa desenvolvido são discutidos.

6.1
Testes de Extração de Atributos Faciais

O ASM treinado foi testado em todas as imagens do banco de imagens CK+ e as formas encontradas foram comparadas com os pontos de referência oficiais disponibilizados em conjunto com o banco de imagens. Para motivos de comparação e escala, é importante ressaltar que as imagens do CK+ possuem resolução de 640x490 *pixels*. O erro quadrático encontrado foi de 40.3 *pixels* para cada forma. Isso indica um erro médio de aproximadamente 0.6 *pixels* para cada ponto de referência. É importante destacar que os pontos de referência oficiais também foram anotados automaticamente com o uso de AAM.

No entanto, esse resultado foi encontrado para o ASM treinado com 90%

PUC-Rio - Certificação Digital Nº 1112647/CB

	90%–10%	2–*fold*
Erro médio por forma	40,3	150,3
Erro médio por ponto	0,6	2,21
Variância da forma	0,011	0,02

Tabela 6.1: Valores de erro médio por forma, ponto e a variância encontrada para o ASM treinado com o banco de imagens CK+ testados em corte de 90% e 10% e também em 2-*fold*

do CK+ e testado com apenas 10% do mesmo. Devido à falta de bancos de imagens que possuem o mesmo padrão de anotações que o CK+, não foi possível fazer o teste em um segundo banco. Foi feita, então, a validação cruzada com a divisão do banco em duas partes iguais, resultando em um erro quadrático médio substancialmente maior, igual a 150.3 *pixels* por forma, 2.21 por ponto.

Um erro médio de 2.21 pixels para cada ponto na forma é satisfatório como etapa inicial do processo de reconhecimento de emoções. Uma validação cruzada com 10 separações traria um resultado mais realista, mas tal teste não foi realizado principalmente por que o erro já se encontrava suficientemente baixo. Além disso, o custo computacional necessário para fazer o treinamento e teste de cada modelo ASM foi um obstáculo para fazer tal teste.

O aspecto fundamental que acarreta um erro médio relativamente baixo é o bom comportamento das imagens. Nas imagens de treinamento o ambiente é completamente controlado. Fatores como a distância da pessoa até a câmera, a pose do indivíduo, a ausência de acessórios como óculos, barba e cabelo sobrepondo partes faciais e, o mais importante, a iluminação, seguem critérios meticulosos durante a sessão de fotos.

A Tabela 6.1 exibe os valores de erro médio encontrados no teste do ASM, incluindo a variância encontrada. Como pode ser visto, a variância do modelo que foi treinado com mais dados é menor do que no feito com validação cruzada, algo já esperado, pois o erro médio é maior.

6.2 Banco de Dados para Classificação de Emoções

Inicialmente, o SVM havia sido treinado com formas extraídas do banco de imagens RaFD, que possui um total de 2241 imagens (após o espelhamento), e a acurácia encontrada foi igual a 91,38% com validação cruzada. No entanto, ao testar esse modelo no banco de dados CK+, percebemos que o modelo estava visivelmente sofrendo com treinamento excessivo do banco RaFD e, portanto, tinha pouca capacidade de generalização, resultando em uma acurácia igual a

PUC-Rio - Certificação Digital Nº 1112647/CB

61,57%.

De tal maneira, optamos por fazer uma mescla dos dois bancos de imagens com a finalidade de criar outro mais genérico, com mais variação étnica e também mais exemplos. O banco de dados final para treinamento é composto pelos exemplos obtidos das imagens do RaFD acrescido de 418 outros providos pelo CK+, totalizando 2659 exemplos.

Um novo modelo foi treinado com 100% do RaFD e 50% do CK+ e testado no restante do CK+. A acurácia para esse modelo e teste foi de 80,18%, indicando que o uso da mescla dos bancos de dados foi produtivo para o modelo.

A partir desse momento, então, tomamos por padrão o uso dos dois bancos de imagens em conjunto para treinamento dos modelos.

6.3 Testes de Classificação de Emoções

Como descrito na seção de Metodologia, os atributos escolhidos para treinamento do modelo de classificação de emoções dependem de dois estados, um inicial e um final. O estado final é aquele que se deseja descobrir, classificar, e o estado inicial é uma forma de referência já conhecida que servirá como base de comparação para extrair o deslocamento dos pontos faciais. Dois modelos SVM foram criados, um que utiliza a face neutra do indivíduo do estado inicial e outro que toma a forma média do ASM como referência. É importante ressaltar que existe a possibilidade de empate de votos em uma predição utilizando o SVM multiclasse. Tal problema não foi tratado diretamente, isto é, deixamos o LIBSVM encarregado pela decisão de desempate.

6.3.1 Face neutra como referência

Essa abordagem tem como vantagem possuir mais informações sobre o indivíduo que está sendo classificado, pois é necessário que a sua face neutra seja conhecida. Como a face neutra é usada como referência se torna inviável a classificação da expressão neutra.

A validação cruzada do modelo em 10 divisões produziu uma acurácia de 87.96%. Mais resultados específicos e a matriz de confusão podem ser vistos nas Tabelas 6.2 e 6.3, respectivamente. Pelos resultados, podemos ver que as classes com pior desempenho foram tristeza e medo. Por outro lado, surpresa e felicidade obtiveram os melhores números.

PUC-Rio - Certificação Digital Nº 1112647/CB

Classe	PV	FP	Precisão	Recall	F-Measure	ROC Area
raiva	0.884	0.025	0.874	0.884	0.879	0.93
desgosto	0.884	0.025	0.879	0.884	0.882	0.93
medo	0.833	0.034	0.814	0.833	0.823	0.899
felicidade	0.936	0.008	0.963	0.936	0.95	0.964
tristeza	0.818	0.034	0.814	0.818	0.816	0.892
surpresa	0.906	0.017	0.916	0.906	0.911	0.944
Média	0.88	0.023	0.88	0.88	0.88	0.928

Tabela 6.2: Resultados encontrados para o SVM com o uso da face neutra como estado inicial de referência, onde PV é positivos verdadeiros e FP falsos positivos

Classe	raiva	desgosto	medo	felicidade	tristeza	surpresa
raiva	381	25	5	3	16	1
desgosto	27	398	6	8	8	3
medo	3	3	338	1	36	25
felicidade	4	13	4	471	8	3
tristeza	17	13	34	5	336	6
surpresa	4	1	28	1	9	415

Tabela 6.3: Matriz de confusão para o SVM com o uso da face neutra como estado inicial de referência

PUC-Rio - Certificação Digital Nº 1112647/CB

6.3.2 Forma média como referência

Essa abordagem toma como estado inicial a forma média do ASM e, nesse momento, ignoramos a sua capacidade de predizer a expressão neutra com a finalidade de promover uma comparação mais precisa entre as duas abordagens para estado inicial. A última seção dos resultados exibe o resultado encontrado para o mesmo modelo com a adição da classe neutra na predição.

O modelo mostrou uma acurácia de 87.1% e maiores detalhes podem ser vistos nas Tabelas 6.4 e 6.5. Assim como no modelo que usa a forma neutra, as classes em que houve maior erro de predição foram tristeza e medo. A classe com melhores resultados continuou sendo a felicidade, já a segunda melhor, dependendo da métrica analisada, pode ser tanto desgosto quanto surpresa.

6.3.3 Comparação entre Forma Média e Neutra

Cada uma dessas abordagens possui seus prós e contras. Para o uso da face neutra como estado inicial, existe a desvantagem de ser necessário ter tal conhecimento *a priori*. Isso não é tão relevante no momento de treinamento, pois existe uma grande disponibilidade de fotos de um mesmo indivíduo nos

Classe	PV	FP	Precisão	Recall	F-Measure	ROC Area
raiva	0.868	0.025	0.87	0.868	0.869	0.921
desgosto	0.891	0.026	0.876	0.891	0.883	0.933
medo	0.771	0.032	0.815	0.771	0.792	0.87
felicidade	0.956	0.012	0.951	0.956	0.953	0.972
tristeza	0.835	0.04	0.792	0.835	0.813	0.897
surpresa	0.88	0.02	0.9	0.88	0.89	0.93
Média	0.871	0.025	0.871	0.871	0.871	0.923

Tabela 6.4: Resultados encontrados para o SVM com o uso da forma média como estado inicial de referência sem classe neutra, onde PV é positivos verdadeiros e FP falsos positivos

Classe	raiva	desgosto	medo	felicidade	tristeza	surpresa
raiva	374	22	4	6	24	1
desgosto	26	401	5	6	10	2
medo	4	9	313	10	40	30
felicidade	3	5	6	481	4	4
tristeza	20	14	23	3	343	8
surpresa	3	7	33	0	12	403

Tabela 6.5: Matriz de confusão para o SVM com o uso da face média como estado inicial desconsiderando a emoção neutra.

PUC-Rio - Certificação Digital Nº 1112647/CB

bancos de imagem. No entanto, do ponto de vista de uma aplicação que usaria esse preditor, seria necessário que o indivíduo tivesse sua face neutra configurada de antemão, pois, sem isso, não seria possível extrair os atributos. Outra desvantagem dessa abordagem é a perda da capacidade de predição do estado neutro. Isso ocorre porque ele é a base da comparação e, portanto, se torna inviável predizê-lo. Como vantagem, o modelo ganha informação sobre a estrutura do rosto do indivíduo, algo que possivelmente aumentaria a acurácia do modelo.

Para a abordagem com o uso da forma média do ASM não é mais necessário ter conhecimentos do indivíduo e, portanto, não existiria mais uma etapa inicial de configuração da face neutra. Isso traz vantagens para a aplicação, no entanto, o modelo perde informações específicas sobre indivíduo. A forma média do ASM, como o próprio nome diz, é uma aproximação dos rostos estudados durante a etapa de treinamento e, portanto, dificilmente proveria mais informação que a face neutra do próprio indivíduo. Por outro lado, como o estado inicial não é mais a face neutra, torna-se possível a predição do estado neutro.

A Tabela 6.6 resume as informações comparativas descritas anteriormente entre as duas abordagens.

	Vantagens	Desvantagens	Estado neutro
Face neutra	Mais informação sobre o indivíduo	Necessidade de conhecer a face neutra do indivíduo *a priori*	Incapaz de predizer
Forma média ASM	Desnecessário configurar a face neutra do indivíduo previamente	Usa uma forma média, menos informação do indivíduo	Capaz de predizer

Tabela 6.6: Resumo das vantagens e desvantagens das abordagens para estado inicial

Ignorando esses fatos e analisando puramente os resultados exibidos nas subseções anteriores, podemos ver que o resultado entre as duas abordagens se mostrou surpreendentemente similar. Enquanto a forma neutra apresentou uma acurácia de 87,96%, a forma média alcançou 87,1%. À primeira vista, seria fácil, porém incorreto, assumir que o modelo com forma neutra alcançaria resultados muito superiores devido ao conhecimento *a priori* embutido com a face neutra. No entanto, a acurácia do modelo em que são desconhecidos o indivíduo e sua face neutra (aquele que usa a forma média) tem uma capacidade quase similar ao que possui mais informações (aquele que usa a forma neutra).

Foi feita, então, uma análise sobre os resultados encontrados e descobrimos que os dois modelos apenas alcançaram uma acurácia próxima devido à quantidade de exemplos contidos no banco de dados. Através do gráfico da Figura 6.1 podemos ver que o SVM treinado com atributos que usaram a face neutra como referência se destacavam mais quando treinado com um banco de dados menor. Quando treinado com aproximadamente 70% do banco de dados, as capacidades de predição dos dois modelos se aproximaram bastante.

6.3.4 Forma média como referência e adição da classe neutra

Esse modelo usa melhor o potencial que a escolha da forma média como estado inicial propicia, isto é, a capacidade de predizer a forma neutra além das outras 6 universais. Devido a esse fator, novos exemplos relacionados à forma neutra surgiram, gerando um banco de dados com 3207 exemplos (838 do CK+ e 2366 do RaFD).

A acurácia desse novo SVM foi de 83.2%. Mais resultados específicos e a matriz de confusão podem ser vistos nas Tabelas 6.7 e 6.8, respectivamente.

PUC-Rio - Certificação Digital Nº 1112647/CB

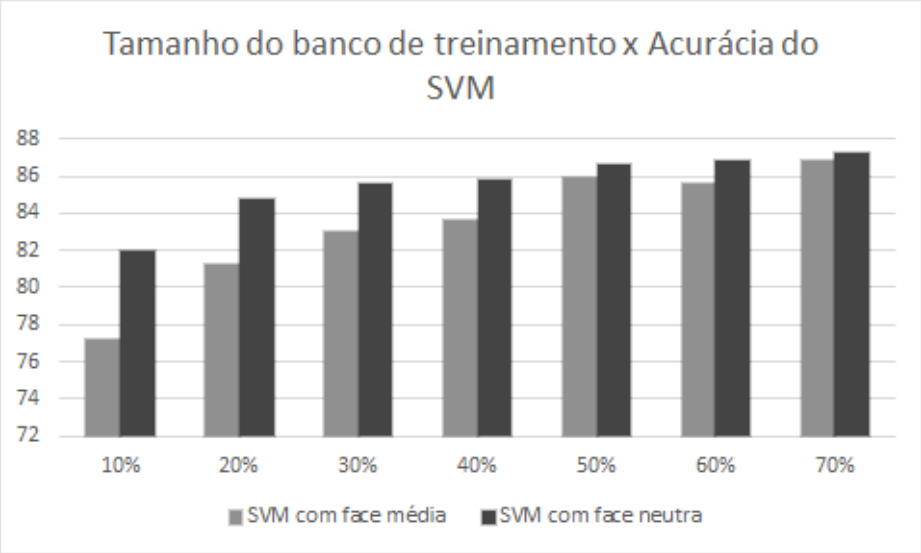


Figura 6.1: Gráfico que exibe a diferença de acurácia dos dois modelos SVM em comparação com o tamanho do banco de dados usado.

Classe	PV	FP	Precisão	Recall	F-Measure	ROC Area
raiva	0.828	0.024	0.844	0.828	0.836	0.902
desgosto	0.891	0.02	0.879	0.891	0.885	0.936
medo	0.764	0.028	0.799	0.764	0.781	0.868
felicidade	0.95	0.011	0.941	0.95	0.946	0.97
neutro	0.791	0.053	0.753	0.791	0.772	0.869
tristeza	0.706	0.044	0.702	0.706	0.704	0.831
surpresa	0.871	0.016	0.899	0.871	0.885	0.927
Média	0.832	0.028	0.833	0.832	0.832	0.902

Tabela 6.7: Resultados encontrados para o SVM com o uso da forma média como estado inicial de referência, onde PV é positivos verdadeiros e FP falsos positivos

Classe	raiva	desgosto	medo	felicidade	neutro	tristeza	surpresa
raiva	357	23	2	3	30	16	0
desgosto	18	401	4	6	12	7	2
medo	4	8	310	9	10	34	31
felicidade	3	4	4	478	7	4	3
neutro	26	6	13	8	431	56	5
tristeza	10	9	21	3	74	290	4
surpresa	5	5	34	1	8	6	399

Tabela 6.8: Matriz de confusão para o SVM com o uso da forma média do ASM como estado inicial de referência

Assim como nos dois modelos anteriores, a classe felicidade apresenta a melhor precisão. No entanto, a classe neutra adicionada também entrou no grupo das classes que estão bem abaixo da média. A classe que mais sofreu

PUC-Rio - Certificação Digital Nº 1112647/CB

com a adição do neutro foi a tristeza, que perdeu 9% de precisão, se tornando, indiscutivelmente, a classe com o pior resultado de predição.

6.3.5 Desempenho

A velocidade final do algoritmo executado em um CPU Intel Core i5-2410M @ 2.30GHz com 6 GB de memória RAM em Windows 7 foi de, em média, 120 ms para a execução do ASM, ou seja, para encontrar os pontos faciais. Após isso, a predição usando o SVM (classificação da emoção com base na forma encontrada) é de 1ms, em média. Claramente, o gargalo se encontra no ASM. Com um total de 121ms uma aplicação seria capaz de executar a aquisição da nova forma e a sua classificação cerca de 8 vezes por segundo, variando entre 5 e 10 vezes por segundo. No entanto, podemos dizer que essa implementação não é otimizada.

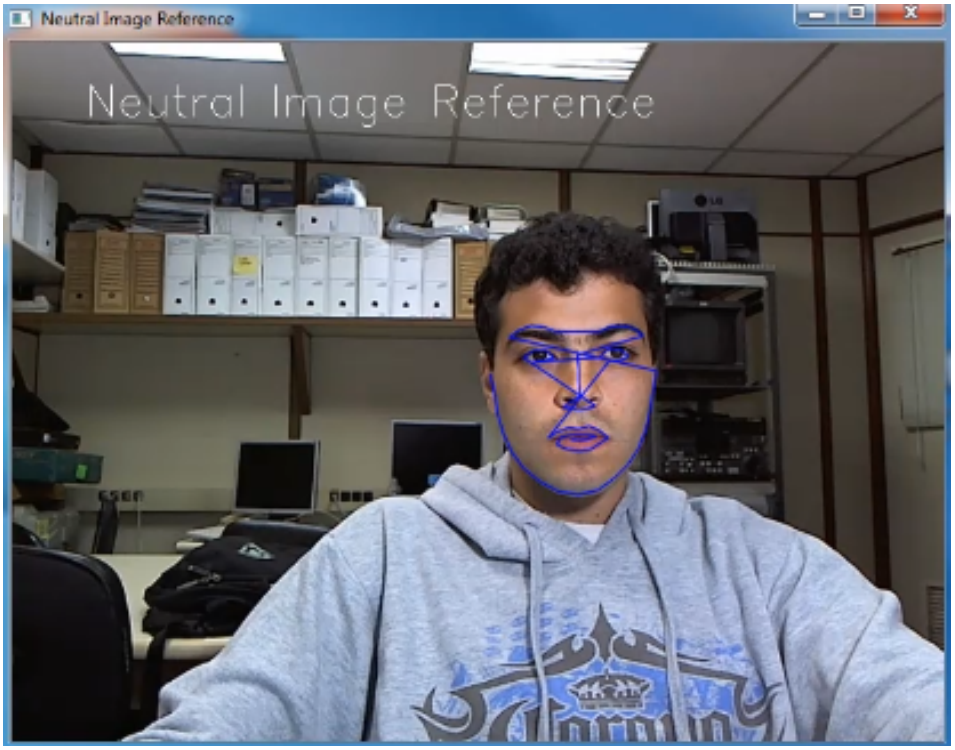
Wang *et al.* (41) alcançaram um tempo de 18.39 ms para o ASM e 0.729 ms para a predição do SVM, registrando um desempenho mais que seis vezes superior ao método final utilizado. Wang *et al.* (*op. cit.*) não deixam explícito decisões de implementação do seu ASM e, portanto, é difícil saber em qual etapa o seu algoritmo ultrapassou o STASM. De qualquer forma, o estudo de Wang *et al.* (*op. cit.*) ainda não é o que mais se destaca em questão de performance. Com o avanço de programação de propósito geral em GPUs, Li *et al.* (2009) utilizando CUDA 2.0, alcançaram um algoritmo que é capaz de aplicar o ASM em 32 imagens em apenas 8.55 ms, ou seja, 3.74 imagens em apenas um milissegundo. A grande diferença do algoritmo de Li *et al.* (*op. cit.*) é que, no momento de refinar a localização dos pontos, a busca é feita em paralelo para cada ponto, o que reduz drasticamente o tempo do ASM, (26) .

6.4 Aplicação para Reconhecimento de Emoções

Um programa foi desenvolvido utilizando os modelos ASM e SVM treinados. Uma imagem da tela do programa pode ser vista na Figura 6.2. A Figura 6.2(a) exibe a imagem neutra de referência e sua forma em azul é aquela que será utilizada como estado inicial no momento da predição. A Figura 6.2(b) apresenta o momento atual de captura, indicando que naquele momento, baseado na imagem neutra definida, o usuário está feliz (levando em consideração a janela deslizante de predições).

Por ter sido utilizado uma janela deslizante de últimas predições, é possível saber a confiança, isto é, probabilidade de uma determinada predição

PUC-Rio - Certificação Digital Nº 1112647/CB



6.2(a): Imagem de referência

6.2(b): Imagem para classificação

Figura 6.2: *Screenshot* do programa, mostrando uma janela com a face de referência 6.2(a) e a imagem para classificação 6.2(b) e a predição no canto superior esquerdo, feliz, no caso.

PUC-Rio - Certificação Digital Nº 1112647/CB

de acordo com o número de ocorrência de cada classe na janela. A Figura 6.3 mostra como, no programa, tal confiança é transmitida para o usuário de maneira gráfica, através de um gráfico de barras. Nesse caso, existe uma certa insegurança entre a emoção triste e a neutra, mas a triste é escolhida por possuir uma maior probabilidade, isto é, a emoção triste é a que possui mais ocorrências na janela deslizante das últimas predições encontradas.

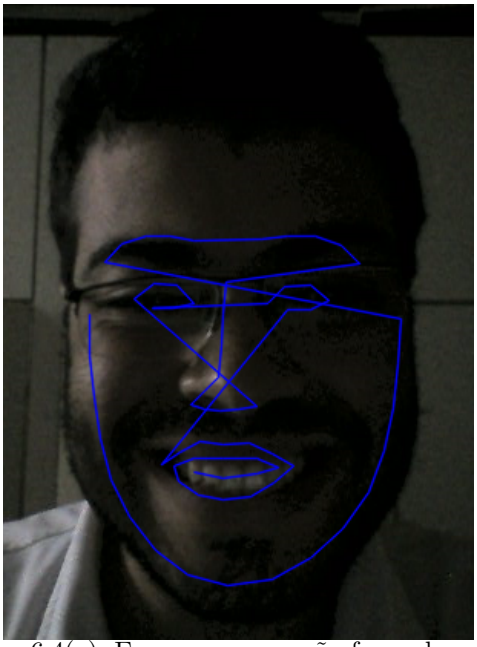
Figura 6.3: Um gráfico de barras exibe a probabilidade das emoções no dado momento.

Pelo fato de que o ambiente e o indivíduo capturados pela aplicação não são necessariamente controlados, a capacidade de extração de pontos faciais se torna menos confiável. Na Figura 6.4 são exibidos exemplos de erros encontrados devido a problemas de iluminação 6.4(a) ou de pose e acessórios 6.4(b).

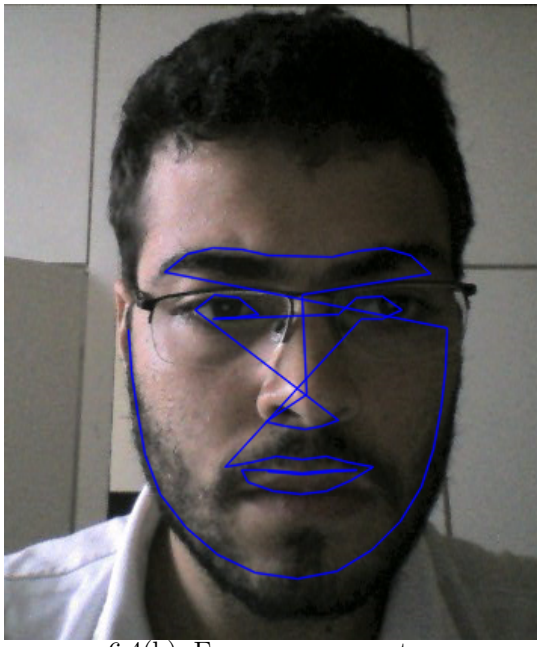
O resultado geral do programa foi que, caso certos requisitos sejam seguidos, a predição do programa se torna satisfatória. Tais requisitos servem para tornar a captura da imagem mais próxima das imagens com as quais o ASM foi treinado, sendo eles:

- O usuário deve estar posicionado frontalmente para a câmera;
- A distância do usuário para a câmera não deve ultrapassar 1 metro;
- A iluminação deve estar presente de forma que ilumine os dois lados do rosto do indivíduo por igual e não dever ser muito forte ou fraca;

PUC-Rio - Certificação Digital Nº 1112647/CB



6.4(a): Erro com expressão formada



6.4(b): Erro em pose neutra

Figura 6.4: Exemplos de erros possíveis devido a pouco controle do ambiente e indivíduo

- Não devem estar presentes acessórios como óculos escuros, barba, cabelo ocluindo a face;
- A câmera deve possuir boa qualidade de captura de imagem.

Como definido inicialmente, era importante que a aplicação executasse em tempo real e a velocidade definida de 5 a 10 *frames* por segundo se demonstrou satisfatória, apesar do atraso de resposta causado pelas janelas deslizantes usadas na aplicação.

PUC-Rio - Certificação Digital Nº 1112647/CB