5
Análise Experimental

5.1.
Base de dados para testes

Foram gravados diversos vídeos que serviram para realizar os testes realizados nesta dissertação. Cada um dos vídeos gerados para medir qualidade da detecção das características da face foi separado em uma seqüência de imagens e cada um dos quadros teve suas características faciais marcadas manualmente. Dessa forma, é possível estimar o erro calculando a diferença entre a posição das características faciais detectadas ao longo do vídeo pelo programa implementado e sua posição real.

Além disso, foram gerados vídeos sintéticos a fim de realizar testes com a implementação original do FERNS. Esses vídeos constam de uma imagem de treinamento projetada no plano e foi implementada uma rotina para gerar, automaticamente, uma seqüência de movimentos idênticos para cada uma das imagens utilizadas. Com isso é possível comparar melhor os resultados, uma vez que a única diferença entre os vídeos é a imagem de treinamento.

Para uma melhor avaliação dos vídeos testados esses foram divididos em quatro tipos de movimento. Considerando o sistema de coordenadas Tait-Bryan angles (figura 12) temos os movimentos divididos em: escala, translação, rotação em torno do eixo longitudinal (roll) e rotação em torno do eixo vertical (yaw). Os movimentos utilizados representam uma simplificação para os movimentos normais do rosto.

Os vídeos utilizados nos testes foram gerados utilizando-se uma Quickcam Pro 9000 da Logitech. Os vídeos foram capturados a uma taxa de 30 quadros por segundo, utilizando uma resolução de 640 x 480.

PUC-Rio - Certificação Digital Nº 0711279/CA

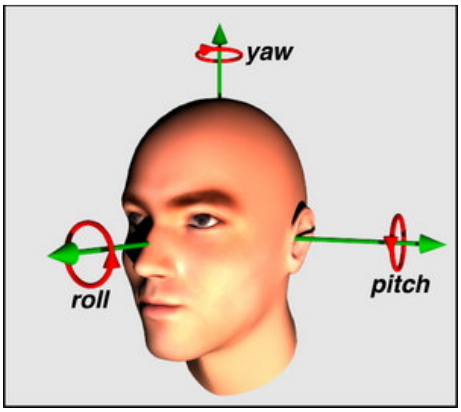


Figura 1: Ângulos de Tait-Bryan (Fonte: www.pittpatt.com).

5.2. Análise do FERNS original

Com o objetivo de avaliar a utilização dos FERNS em imagens de face, três tipos de testes foram realizados. O primeiro teste examina a importância da textura para uma correta detecção usando o FERNS. Além disso, verifica se a textura das faces é suficiente para se obter bons resultados. O segundo conjunto de testes estuda a influência dos tamanhos das regiões a serem detectadas. Finalizando, o terceiro tipo de teste analisa o uso da planaridade para geração das vistas sintéticas no caso das faces.

5.2.1. Influência da textura na detecção pelo FERNS

Para analisar a influência das texturas na detecção pelo FERNS foram gerados dois vídeos sintéticos. O primeiro vídeo utiliza a imagem da face sem modificações, enquanto no segundo vídeo foram adicionados retalhos de textura à face, como mostrado na figura 13. Nessa figura, o quadrado desenhado ao redor da face representa a área a ser detectada pelo classificador.

Projetando cada uma das imagens em um plano, foram simulados movimentos de translação, escala, rotação em torno do eixo longitudinal (roll) e rotação em torno do eixo vertical (yaw). Exemplos desses movimentos podem ser visto na figura 14. Os movimentos realizados em ambos os vídeos são idênticos, de forma a possibilitar uma comparação entre eles.

PUC-Rio - Certificação Digital Nº 0711279/CA

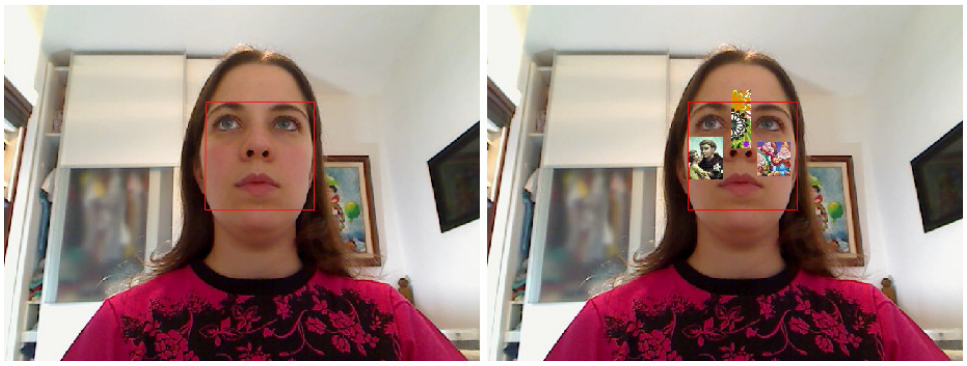


Figura 2: (a) Imagem da face. (b) Imagem da face com adição de texturas.

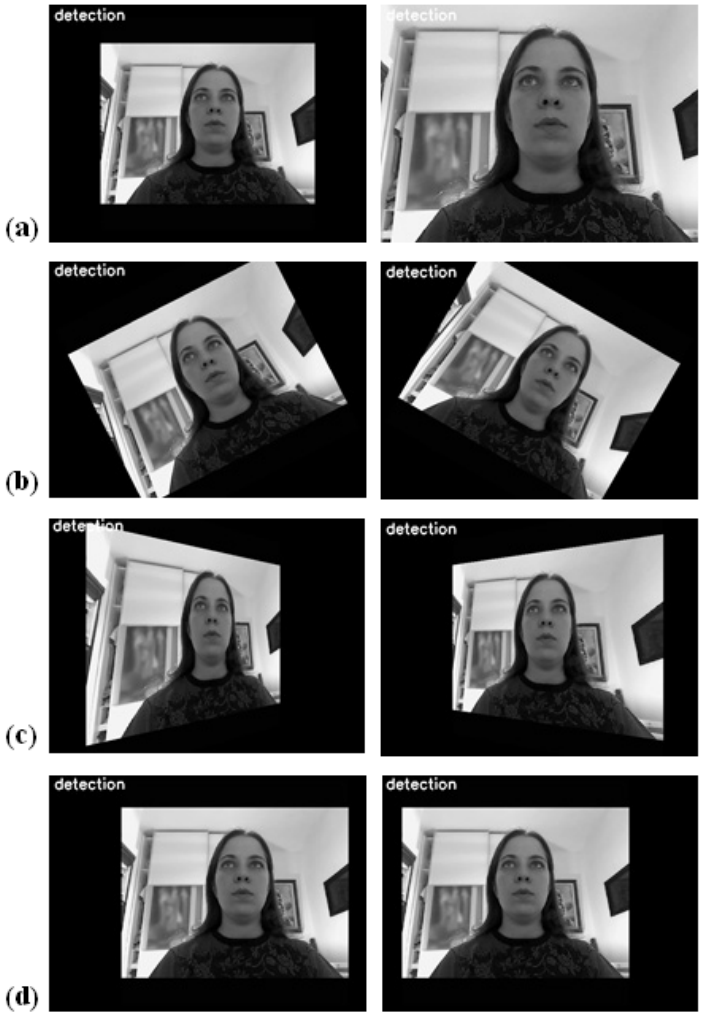


Figura 3: Movimentos simulados nos vídeos sintéticos: (a) escala, (b) rotação em relação ao eixo longitudinal (roll), (c) rotação em relação ao eixo vertical (yaw) e (d) translação.

PUC-Rio - Certificação Digital Nº 0711279/CA

O classificador obtido através do treinamento da imagem da face sem modificações é utilizado para a detecção da face no vídeo construído a partir dessa imagem. Da mesma forma, o classificador gerado a partir da imagem com retalhos é utilizado na detecção da face no vídeo construído a partir da imagem com retalhos de textura. Com isso é possível calcular a taxa de detecção correta em cada um dos vídeos. Para o vídeo sintético sem textura a taxa de detecção foi de 57,7%, enquanto para o vídeo com textura esse valor é de 97,7%. Isso demonstra a importância de se ter uma imagem bem texturizada para o bom funcionamento do FERNS.

Para verificar se a textura das características da face é suficiente para se obter bons resultados foram colocados retalhos de textura em cima da região dos olhos e das narinas, como mostrado na figura 15.

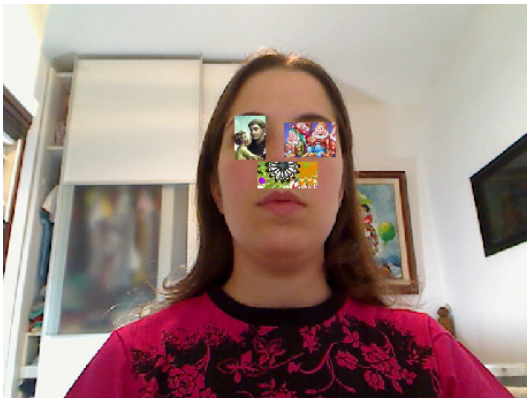


Figura 4: Imagem da face com textura sobre as características faciais.

A taxa de reconhecimento obtida com a imagem com retalhos de textura sobre os olhos e narinas foi de 94.1%. Com isso é possível afirmar que, utilizando os FERNS, as características faciais não são suficientes para garantir uma alta taxa de detecção, já que quando essas foram substituídas por texturas melhores (mais detalhadas e não simétricas) houve um grande aumento na quantidade de quadros reconhecidos corretamente. O que não aconteceu quando a imagem não possuía as texturas colocadas artificialmente e obteve apena 57,7% de detecções corretas. Os resultados obtidos nos testes desta seção estão sintetizados na tabela 1.

PUC-Rio - Certificação Digital Nº 0711279/CA

Face	57.7%
Face com textura e características faciais	97.7%
Face com textura sobre as características faciais	94.1%

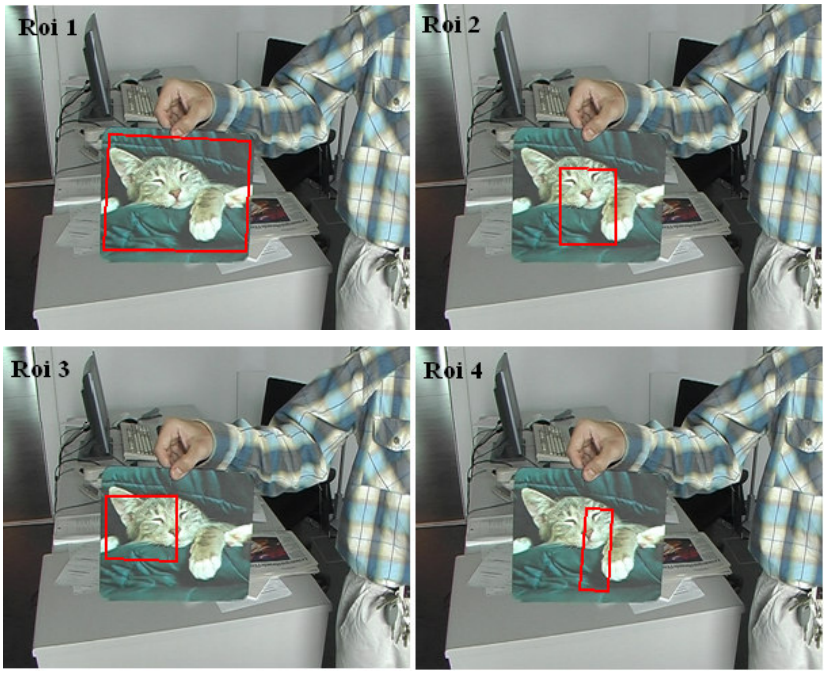
Tabela 1: Taxas de detecções corretas

5.2.2. Influência do tamanho da região de interesse na detecção pelo FERNS

Para analisar a relação do tamanho das regiões a serem detectadas com a taxa de detecção correta, dois vídeos foram gerados utilizando duas imagens de treinamento diferentes.

EXEMPLO 1:

As regiões de interesse (ROI) utilizadas para os testes com o primeiro vídeo podem ser vistas na figura 16. A tabela 2 mostra as taxas de detecção obtidas para cada região de interesse desse exemplo.



PUC-Rio - Certificação Digital Nº 0711279/CA

Figura 5: Áreas de interesse do exemplo 1.

Vídeo 1	Tamanho (w x h)	Área (pixels)	Taxa de detecção (%)
ROI 1	255 x 200	51000	86.6
ROI 2	102 x 136	13872	39.0
ROI 3	122 x 101	12322	26.8
ROI 4	48 x 147	7056	0.0
ROI 5	60 x 43	2580	0.0
ROI 6	35 x 73	255	0.0

Tabela 2: Taxas de detecção obtidas para cada Roi.

EXEMPLO 2:

As regiões de interesse (ROI) utilizadas para os testes com o segundo vídeo podem ser vistas na figura 17. A tabela 3 mostra as taxas de detecção obtidas para esse exemplo.

PUC-Rio - Certificação Digital Nº 0711279/CA

Figura 6: Áreas de interesse do exemplo 2.

Vídeo 1	Tamanho (w x h)	Área (pixels)	Taxa de detecção (%)
ROI 1	337 x 260	87620	96.8
ROI 2	159 x 96	15264	26.6
ROI 3	82 x 57	4674	0.6
ROI 4	32 x 124	3968	0.0
ROI 5	82 x 35	2870	0.0
ROI 6	46 x 43	1978	0.0

Tabela 3: Taxas de detecção obtidas para cada ROI.

Observando os resultados obtidos nos exemplos 1 e 2 acima, podemos perceber que quanto menor a área de interesse menor é a taxa de detecção. Isso acontece mesmo em imagens muito texturizadas, como é o caso da imagem utilizada no exemplo 2. Com isso podemos concluir que a detecção de características faciais fica afetada por possuírem uma área de interesse pequena.

PUC-Rio - Certificação Digital Nº 0711279/CA

5.2.3.
Testando a planaridade para faces

Como descrito em [17] e já menciado na seção 4.2, se o objeto possui uma pequena curvatura uniforme, como no caso da face, é possível assumir que os keypoints são localmente planares e suas distorções podem ser vistas como homografias. Özuysal et al. [18] testaram o FERNS em imagens de face assumindo sua planaridade para geração das vistas sintéticas e, de acordo com os autores, bons resultados foram encontrados.

Para testar o uso da planaridade para geração das vistas sintéticas no caso de faces foram criados, primeiramente, dois vídeos. Um dos vídeos consiste da imagem de treinamento projetada em um plano, onde os movimentos do plano simulam os movimentos de escala, translação, rotação em torno do eixo longitudinal e rotação em torno do eixo vertical da mesma forma que explicado na seção 5.2.2. Dessa maneira, a imagem não sofre deformações mesmo com mudanças de perspectiva. O segundo vídeo consiste de uma sequência de imagens da face no qual ocorrem variaçoes da imagem devido às mudanças de perspectiva e também devido a oclusões. Neste trabalho considera-se que não ocorram grandes oclusões, tais como, desaparecimento completo de uma das características faciais. Utilizando a mesma imagem de treinamento para ambos os vídeos e a mesma região de interesse, foi realizado o treinamento dos FERNS e após isso o classificador treinado foi utilizado para detecção da região de interesse nos demais quadros de cada um dos vídeos. A imagem de treinamento utilizada no teste e alguns quadros detectados, em cada um dos casos podem ser vistos na figura 18.

Além dos dois vídeos utilizados no teste acima, mais dois outros vídeos foram gerados da mesma forma que explicado no parágrafo anterior. Uma nova imagem de treinamento foi utilizada e foram medidas as taxas de detecção para o caso do vídeo com o plano e para o vídeo da face. A figura 19 mostra a imagem de treinamento utilizada. As taxas de detecção para o primeiro e segundo exemplos podem ser vistos na tabela 4.

Os resultados obtidos nos dois exemplos foram semelhantes tanto para o vídeo da face quanto para o vídeo da imagem de treinamento projetada no plano. No entanto, ao contrário do que foi falado no trabalho [18], não foram obtidos

PUC-Rio - Certificação Digital Nº 0711279/CA

resultados satisfatórios para o caso das faces utilizando-se a implementação original do FERNS.

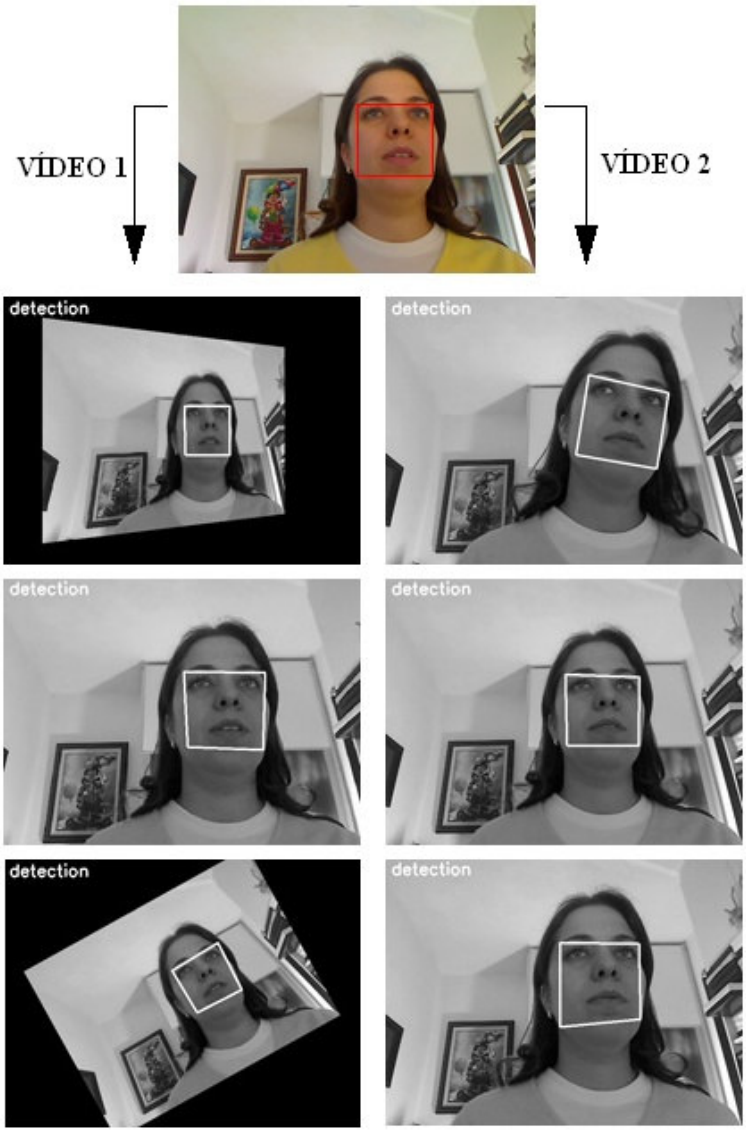


Figura 7: Exemplo 1 - Imagem de treinamento e exemplos de quadros detectados: a primeira coluna representa os quadros detectados no vídeo da face projetada no plano e a segunda coluna representa quadros detectados no vídeo da face.

PUC-Rio - Certificação Digital Nº 0711279/CA

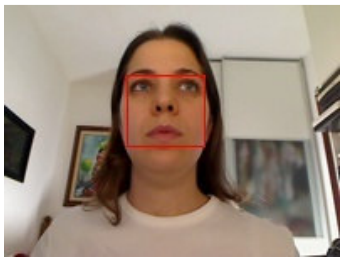


Figura 8: Imagem de treinamento utilizada no exemplo 2.

Vídeo	Taxa de detecção (%)
Exemplo 1 – face	15.9
Exemplo 1 – face no plano	14.4
Exemplo 2 – face	35.1
Exemplo 2 – face no plano	27.5

Tabela 4: Taxas de detecção corretas obtidas para cada vídeo, utilizando o FERNS original.

5.3. Clusterização dos keypoints reconhecidos

Como mostrado na seção 5.2, a face não é suficientemente texturizada de forma a obter altas taxas de detecção utilizando os FERNS. A solução proposta nessa dissertação para resolver o problema foi utilizar grupos de keypoints próximos. Como explicado anteriormente (seção 4.5), duas abordagens foram implementadas para encontrar as regiões das características da face. Na primeira abordagem implementada, para cada uma das características da face, considera-se o maior grupo como sendo a região onde essa se encontra. No segundo método considera-se o cluster mais forte, isto é, o cluster cuja soma das probabilidades é a maior. Nesta seção é realizada uma comparação entre esses dois métodos. Para isso, quatro vídeos com imagens de faces foram gerados. Para analisar melhor os resultados esses vídeos foram separados em movimentos de escala, translação rotação em torno do eixo longitudinal e rotação em torno do eixo vertical. O primeiro quadro de cada vídeo foi utilizado como imagem de treinamento e podem ser vistos na figura 20.

PUC-Rio - Certificação Digital Nº 0711279/CA

Também foi considerada nos testes a distância entre os keypoints de um cluster (agrupamento). As tolerâncias para as distâncias entre os pontos dos clusters considerados foram de 10 e 15 pixels.

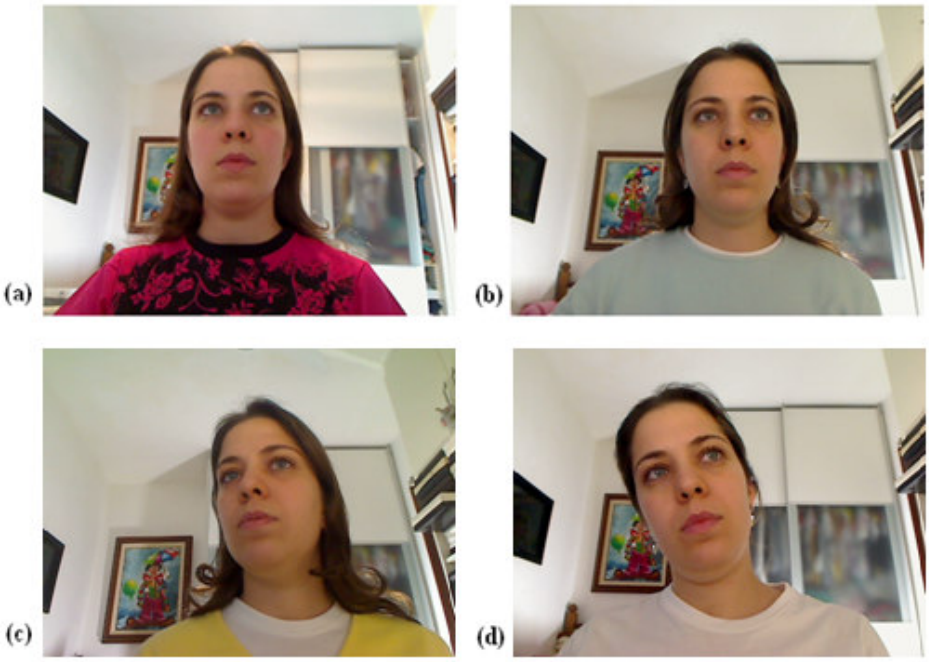


Figura 9: Imagem de treinamento dos vídeos: (a) escala, (b) translação, (c) rotação em torno do eixo vetical e (d) rotação em torno do eixo longitudinal.

As características faciais levadas em consideração nos testes foram: narina direita (ND), narina esquerda (NE), olho direito e esquerdo representados pelo centro da elipse ajustada à região do olho (ODC e OEC, respectivamente) e olho direito e esquerdo representados pela coluna de pixels pretos no interior da elipse ajustada à região do olho (ODM, OEM). Ambos os métodos para detecção da pupila foram explicados na seção 4.7.2.

A tabela 5 apresenta as taxas de detecções corretas para cada característica da face, onde primeiramente considera-se correto erros menores que 6 pixels e após isso, consideram-se a porcentagem de detecções onde o erro é menor do que 10 pixels. A figura 21 mostra exemplos de quadros detectados dentro de cada uma dessas faixas de valores, deixando claro que erros dessa magnitude ainda são considerados aceitáveis. Isso se deve também ao fato de que as marcações manuais para cálculo de erros não tão precisas, podendo apresentar certos deslocamentos em relação a área exata do centro da pupila. A primeira parte da

PUC-Rio - Certificação Digital Nº 0711279/CA

tabela 5 apresenta os resultados obtidos utilizando-se o método do cluster mais forte, enquanto a segunda parte utiliza a abordagem de maior cluster.

Cluster mais forte		Erro < 6 pixels (%)				Erro < 10 pixels (%)			
		Escala	**Transl.**	**Yaw**	**Roll**	**Escala**	**Transl.**	**Yaw**	**Roll**
Tolerância = 10	ND	100.0	87.5	100.0	94.0	100.0	88.1	100.0	94.0
	NE	100.0	87.5	100.0	94.0	100.0	88.1	100.0	94.0
	ODC	57.4	81.9	71.9	76.4	99.2	99.4	90.2	84.7
	ODM	14.0	10.0	52.2	42.5	73.4	33.1	72.3	53.5
	OEC	89.0	57.5	77.7	76.4	99.0	95.6	79.9	85.7
	OEM	8.6	11.3	37.1	22.3	69.6	56.3	70.5	39.2
Tolerância = 15	ND	100.0	87.5	99.1	94.0	100.0	88.1	99.1	94.0
	NE	100.0	87.5	99.1	94.0	100.0	88.1	99.1	94.0
	ODC	56.6	81.9	65.6	77.1	99.4	99.4	83.9	84.7
	ODM	12.6	11.9	51.3	46.2	76.4	30.6	71.9	55.8
	OEC	90.6	60.0	78.6	79.7	100.0	95.6	80.8	90.4
	OEM	12.2	15.0	40.6	23.6	72.6	67.5	70.1	40.5

Maior Cluster		Erro < 6 pixels (%)				Erro < 10 pixels (%)			
		Escala	**Transl.**	**Yaw**	**Roll**	**Escala**	**Transl.**	**Yaw**	**Roll**
Tolerância = 10	ND	100.0	87.5	100.0	91.0	100.0	88.1	100.0	91.0
	NE	100.0	87.5	100.0	91.0	100.0	88.1	100.0	91.0
	ODC	54.0	83.1	72.8	75.7	97.2	100.0	91.1	83.1
	ODM	13.2	10.6	52.7	42.5	75.8	34.4	72.8	52.8
	OEC	86.4	58.8	76.3	78.4	96	96.9	77.2	88.4
	OEM	7.6	11.3	35.7	22.9	67.2	60.0	68.8	40.5
Tolerância = 15	ND	100.0	87.5	99.1	91.7	100.0	88.1	99.1	91.7
	NE	100.0	87.5	99.1	91.7	100.0	88.1	99.1	91.7
	ODC	55.0	81.9	67.0	77.4	97.2	100.0	85.3	83.7
	ODM	12.4	11.9	51.3	44.2	75.8	32.5	71.9	53.8
	OEC	90.0	60.6	75.4	80.4	99.0	96.9	76.8	93.0
	OEM	12.2	15.0	39.3	24.6	72.2	67.5	67.0	41.7

Tabela 5: Taxas de detecções corretas.

PUC-Rio - Certificação Digital Nº 0711279/CA

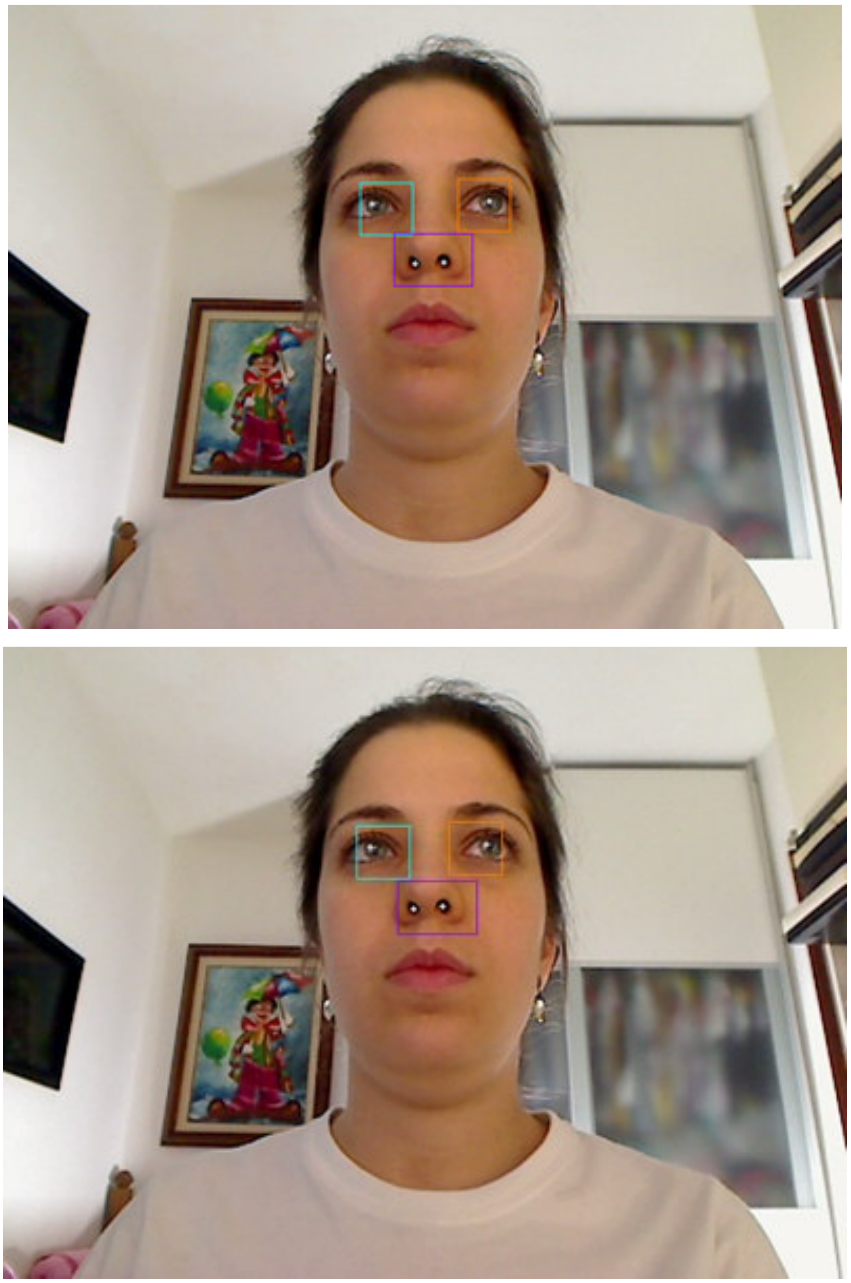


Figura 10: Valores de erro encontrados. A imagem superior apresenta erros iguais a: ND = 1.0, ND = 2.0, OD = 4.2 e OE = 2.2. A imagem inferior apresenta erros iguais a: ND = 1.0, NE = 2.0, OD = 5.0 e OE = 5.0.

5.4. Rastreamento

Para testar o rastreamento o programa implementado foi executado usando, para comparação, a abordagem de cluster mais forte sem levar em consideração a detecção do quadro anterior. Em seguida, na execução do programa considerou-se

PUC-Rio - Certificação Digital Nº 0711279/CA

a detecção obtida no quadro anterior, isto é, foi usado o rastreamento. A distância utilizada para se ativar o rastreamento foi de 40 pixels, ou seja, se o centro da região de busca do quadro atual se encontrar a uma distância de 40 ou mais pixels do centro da região de busca do quadro anterior, então a região de busca do quadro anterior é mantida. Os resultados obtidos com e sem o rastreamento podem ser vistos na tabela 6.

Rastreamento	Erro < 6 pixels (%) → Tolerância = 15			
	Escala	Transl.	Yaw	Roll
ND	100.0	87.5	100.0	100.0
NE	100.0	87.5	100.0	99.7
ODC	55.8	83.1	79.5	88.4
ODM	12.4	12.5	54.5	47.5
OEC	91.0	61.3	86.6	81.4
OEM	12.2	15.0	47.3	25.3

Cluster mais forte	Erro < 6 pixels (%) → Tolerância = 15			
	Escala	Transl.	Yaw	Roll
ND	100.0	87.5	99.1	94.0
NE	100.0	87.5	99.1	94.0
ODC	56.6	81.9	65.6	77.1
ODM	12.6	11.9	51.3	46.2
OEC	90.6	60.0	78.6	79.7
OEM	12.2	15.0	40.6	23.6

Tabela 6: Taxas de detecção com e sem rastreamento.

Em alguns casos os resultados com e sem rastreamento não apresentaram grande diferença, como no caso do vídeo da escala. Isso acontece por que, na grande maioria dos quadros, a distância entre o centro das regiões de busca de dois quadros consecutivos não foi superior a 40 pixels. O gráfico da figura 22 mostra que no caso do vídeo da escala ocorre apenas um pico de erro e esse pico desaparece quando o rastreamento está ativo.

PUC-Rio - Certificação Digital Nº 0711279/CA

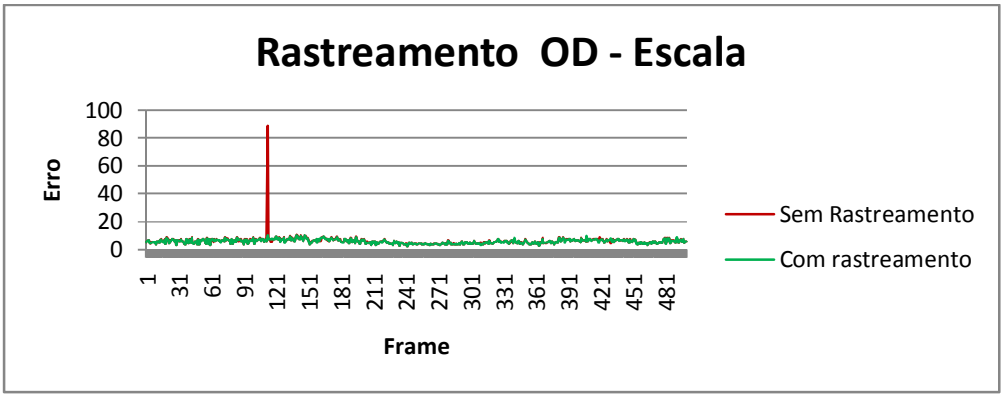


Figura 11: Gráfico de erro do olho direito com e sem rastreamento para o vídeo de escala.

No caso do vídeo da rotação sobre o eixo vertical (yaw), quando o rastreamento está ativo a melhora na taxa de detecção é visível. Isso se dá por que sem o rastreamento muitos erros acontecem, como pode ser visto na figura 23. Com o rastreamento esses picos de erro diminuíram, pois foi mantida a região de busca do quadro anterior. Quando a região de busca é mantida ela ainda se encontra próxima a característica a ser detectada, pois a diferença entre dois quadros consecutivos é muito pequena. Dessa maneira, a característica ainda pode ser reconhecida de forma correta.

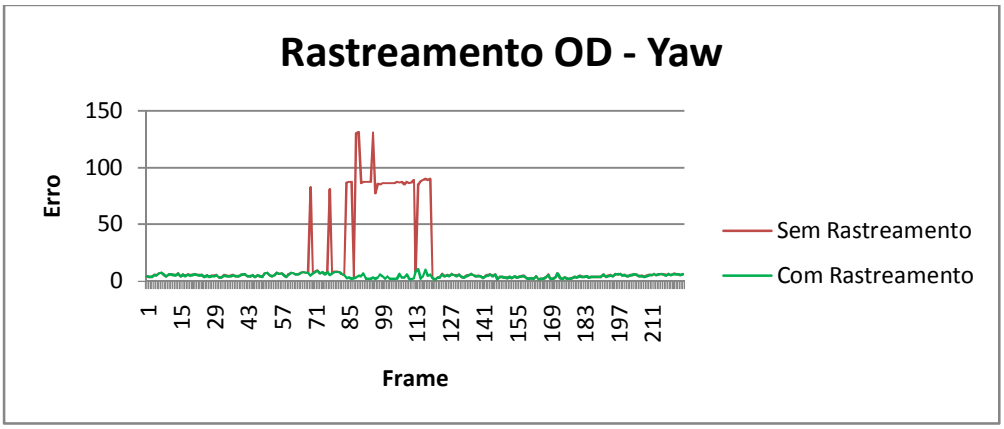


Figura 12: Gráfico de erro do olho direito com e sem rastreamento para o vídeo de rotação sobre o eixo do pescoço.

PUC-Rio - Certificação Digital Nº 0711279/CA

5.5.
Testes com pessoas diferentes

Com o objetivo de melhor validar os resultados obtidos foram realizados testes com pessoas diferentes, apresentando variação na cor da pele e dos olhos. Nesses testes foram medidas as taxas de detecção corretas (erros menores que seis pixels), além da média do erro obtido ao longo dos vídeos para cada uma das características faciais.

EXEMPLO 1:

No primeiro exemplo foram gerados quatro vídeos, cada um representando um tipo diferente de movimento (escala, translação, rotação sobre o eixo longitudinal e rotação sobre o eixo vertical). A imagem de treinamento utilizada para os testes do exemplo 1 pode ser vista na figura 24. A tabela 7 apresenta as taxas de detecção corretas, seguida pela tabela 8 que apresenta o erro médio de cada característica ao longo de cada vídeo testado.

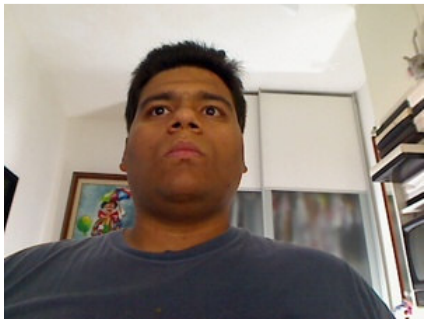


Figura 13: Imagem de treinamento utilizada no exemplo 1.

Vídeo 1	Escala	Translação	Yaw	Roll
ND	93.9	98.4	95.2	93.9
NE	93.1	88.0	89.6	91.2
ODC	86.7	82.5	61.0	95.7
ODM	80.1	64.9	80.1	89.0
OEC	85.9	87.3	96.0	98.8
OEM	67.9	76.9	93.6	86.9

Tabela 7: Taxas de detecção com erro < 6 pixels.

PUC-Rio - Certificação Digital Nº 0711279/CA

Vídeo 1	Escala	Translação	Yaw	Roll
ND	3.6	2.6	3.8	3.8
NE	3.5	3.9	3.8	3.1
ODC	3.9	3.6	4.6	3.6
ODM	5.7	10.0	5.2	4.2
OEC	4.2	3.5	3.3	3.6
OEM	4.5	4.2	3.5	3.6

Tabela 8: Erro médio em pixels

EXEMPLO 2:

Assim como no exemplo 1, no segundo exemplo foram gerados quatro vídeos, cada um representando um tipo diferente de movimento (escala, translação, rotação sobre o eixo longitudinal e rotação sobre o eixo vertical). A imagem de treinamento utilizada para os testes do exemplo 2 pode ser vista na figura 25. A tabela 9 apresenta as taxas de detecção corretas, seguida pela tabela 10 que apresenta o erro médio de cada característica ao longo de cada vídeo testado.

Figura 14: Imagem de treinamento utilizada no exemplo 2.

PUC-Rio - Certificação Digital Nº 0711279/CA

Vídeo 1	Escala	Translação	Yaw	Roll
ND	100.0	99.4	99.0	100.0
NE	100.0	100.0	100.0	100.0
ODC	93.3	79.6	76.0	83.4
ODM	86.3	56.9	50.3	72.4
OEC	100.0	93.4	99.0	97.8
OEM	82.1	76.8	83.4	77.3

Tabela 9: Taxas de detecção com erro < 6 pixels.

Vídeo 1	Escala	Translação	Yaw	Roll
ND	1.4	1.3	1.4	1.4
NE	0.9	1.1	0.8	0.9
ODC	4.2	3.8	4.6	4.9
ODM	4.5	5.3	5.6	5.2
OEC	1.4	2.8	2.0	3.1
OEM	3.5	4.8	3.7	5.3

Tabela 10: Erro médio em pixels para cada característica facial.

PUC-Rio - Certificação Digital Nº 0711279/CA

5.6. Desempenho

Para os testes de performance do método proposto, a linguagem de programação utilizada foi C++. A configuração do computador utilizado nos testes é: Intel Core 2 Quad CPU Q6600, 2.4GHz com 3 GB de RAM. A resolução dos vídeos utilizados é 640x480.

O tempo necessário para cada quadro foi, em média, de 30 ms. Nesse tempo está incluso abertura da imagem e a detecção das características faciais. Como o tempo máximo para considerar uma aplicação tempo real são 33 ms, ou seja, a uma taxa de 30 fps, podemos considerar que o método proposto funciona em tempo real. O tempo médio de processamento dos quadros para cada um dos vídeos da seção anterior pode ser vista na tabela abaixo.

Vídeo	ms
Escala	29.14
Translação	28.36
Yaw	29.03
Roll	29.98

Tabela 11: Tempo médio necessário para o processamento de cada quadro.

PUC-Rio - Certificação Digital Nº 0711279/CA