

## 7

### Conclusão e Trabalhos Futuros

As expressões faciais trazem informações muito importantes que podem auxiliar e melhorar a interação entre homens e computadores. Tais informações podem ser úteis para TV/Cinema e *storytelling* interativo, jogos digitais e robótica. Em vista disso, o trabalho desenvolvido buscou fazer o reconhecimento automático de emoções em tempo real baseando-se em fotos de indivíduos. Para tal, os métodos utilizados foram o de *Active Shape Model* (ASM) (8) para a extração de pontos faciais de imagens e *Support Vector Machines* (SVM) (11) para reconhecimento de padrões das formas encontradas. Não se levou em consideração, neste trabalho, a captura e reconhecimento de micro-expressões.

O ASM foi treinado com o banco de imagens CK+ (22) (27). As partes selecionadas para treinamento foram espelhadas para aumentar o número de exemplos disponíveis. Tal modelo se mostrou satisfatório para extrair pontos faciais de fotos com ambiente controlado, assim como as presentes no CK+.

Utilizamos uma mescla de dois bancos de imagens para treinamento do SVM, RaFD (24) e CK+, o que aumentou a capacidade de generalização dos SVMs por incluir não apenas mais exemplos mas também maior variação étnica e de traços faciais.

Os atributos classificados pelo SVM consistiam no deslocamento entre duas formas, uma representando um estado de referência e a outra, o estado que se desejava reconhecer. Para o estado de referência, duas abordagens foram utilizadas: a forma média do ASM, visto em Wang *et al.* (2010) (41), e a forma neutra do indivíduo, visto em Gang *et al.* (2009) (15), resultando em dois SVMs distintos. Tais modelos foram treinados utilizando a mesma base de dados e imagens, o mesmo algoritmo de extração de pontos e treinamento e também as mesmas classes de emoções, o que possibilitou um ambiente homogêneo para comparação.

Percebemos que apesar do SVM que utiliza a forma neutra possuir informações adicionais sobre o indivíduo que está sendo analisado, quando o número de exemplos para treinamento se torna suficientemente grande, o SVM

que utiliza a forma média como referência consegue resultados muito similares, cerca de 87% de acurácia. Além disso, para os dois modelos as classes em que o preditor mais falhou foram tristeza e medo.

Como os resultados dos dois modelos se mostraram muito próximos, o uso da face média como referência se torna bastante interessante para aplicações que desconhecem o usuário que será analisado, sem perder muita acurácia com isso. No entanto, caso o indivíduo possua traços faciais muito distintos é possível que o sistema falhe completamente para a classificação de todas as emoções, algo que poderia ser contornado com a calibração da sua face neutra *a priori*, utilizando a outra abordagem. A decisão de qual caminho seguir para estado de referência é dependente da aplicação, mas é importante ressaltar que, após a análise feita nessa dissertação, sabemos que não existe uma diferença muito grande de acurácia entre os dois modelos, comprovado pelos testes feitos.

Além dos dois SVMs citados anteriormente, foi criado um terceiro que usa a face média como estado de referência e com adição da classe neutra para a predição. Alcançamos uma acurácia de aproximadamente 82% para esse último modelo.

O desempenho alcançado para a extração de atributos e classificação foi de 121ms em média e um programa foi feito para verificar o uso do método proposto em uma aplicação real. O *loop* principal do programa que consistia em analisar a imagem capturada pela *webcam*, extrair os atributos faciais, classificá-los e exibir a classe encontrada para o usuário era executado em uma média de 5 a 10 vezes por segundo. Na prática, percebemos que o ASM não era tão estável quanto esperado e, por isso, a capacidade de predição era insatisfatória. Alguns ajustes foram feitos para reverter tal situação, com destaque para janelas deslizantes de média para últimas formas encontradas e últimas predições feitas para suavização de erros e também a ampliação de expressões. Os ajustes melhoraram a aplicação no geral, mas vários problemas relacionados à extração de pontos faciais permaneceram.

Pela aplicação desenvolvida foi possível entender o quão importante era o controle do ambiente e do indivíduo no momento de realizar a captura da imagem. O ASM demonstrava uma capacidade de convergência muito baixa quando certos padrões não eram seguidos, como uma boa iluminação e a ausência de acessórios.

O trabalho possui vários caminhos para melhoria, sendo eles principalmente na etapa de extração de pontos faciais e aprendizado de máquina. Para a extração, o desempenho poderia ser melhorado com o uso de algoritmos mais ágeis ou até mesmo através de programação em GPU, como no estudo de Li

*et al.* (26). A melhoria do desempenho possibilitaria até mesmo o uso desse método para vários indivíduos em uma mesma imagem, algo que poderia ser bastante interessante para mídias interativas, como TV digital. Além disso, o treinamento do ASM com um banco de imagens mais extenso e com menos controle de iluminação e de pose do indivíduo favoreceria uma aplicação que tem como objetivo reconhecer esse tipo de pessoa. Para elevar a qualidade dos testes de acurácia da extração de pontos do ASM, seria interessante usar uma métrica de normalização para que os resultados se tornassem independentes da resolução da imagem, como a distância intra-ocular, por exemplo.

Seria interessante, também, fazer testes de acurácia com o programa, isto é, selecionar vários indivíduos, coletar vídeos e fazer uma análise mais precisa de como o método proposto se comporta na prática. Para isso seria necessário desenvolver um banco de imagens com um padrão de controle do ambiente e indivíduo, ainda que bem minimalista e exercer o mesmo padrão durante a execução do programa.

Mudanças nos pilares da dissertação também podem ser feitas, como o uso de dados em 3D extraídos com sensores de profundidade e Hidden Markov Model como algoritmo de aprendizado de máquina para aproveitar sequências de imagens, fazendo com que estados anteriores influenciem na predição.