Relatório do Projeto

Computação Multimédia

Introdução

Este projeto tem como principal objetivo reproduzir um vídeo adequado ao ambiente em que os utilizadores se inserem, baseado no reconhecimento facial e corporal, captados na camara.

O sistema também oferece uma interface para o administrador, sendo possível este definir quando uma determinada imagem ou vídeo deve ser mostrado.

State of Art

O nosso projeto consiste em produzir um vídeo baseado nas reações do utilizador e no ambiente à volta dele. Mais especificamente, o conteúdo que vai aparecendo no vídeo depende da possível quantidade de pessoas, assim como gestos.

Um dos principais recursos é o reconhecimento facial, utilizado para determinar o número de utilizadores assim como os seus movimentos.

Para além desta reprodução do vídeo, o sistema tem uma interface administrador, onde é permitido associar mais especificamente o conteúdo a ser mostrado. Nessa interface de administrador, vamos ter dois tipos de galeria: imagens e vídeos, onde os metadados estão associados. Esses são armazenados em ficheiros XML para evitar a repetição exaustiva de extração de metadados, quer de imagens, quer de vídeos. Este processo é semiautomático, pois depende de um algoritmo de similaridade de vídeo ou imagens.

Para detetar as possíveis similaridades do conteúdo, os metadados só são processados quando um vídeo ou imagem é carregada.

Algoritmos e Técnicas

Os metadados exigidos para cada imagem e vídeo são os seguintes:

- Tags
- Evento
- Cor
- Luminância
- Número de rostos que aparecem na imagem/vídeo
- Edge Distribution
- Características da textura
- Número de vezes que um objeto específico aparece
- Ritmo

Vamos agora explicar os algoritmos e as diferentes técnicas que usamos para analisar estes metadados.

1. Tags

Esta componente representa o conteúdo da imagem/vídeo, tal como tema, local, entre outros.

As diferentes tags de uma imagem/vídeo estão descritos no nome da própria, sendo retirados só a parte que de facto tem as tags.

As tags caracterizam a imagem/vídeo, ou seja, se uma imagem representar Lisboa, uma das tags há de ser Lisboa.

2. Evento

Esta componente representa o momento, de acordo com o utilizador, em que a dada imagem/vídeo será mostrado.

O evento está descrito no nome da imagem/vídeo, sendo, tal como nas tags, retirado, só a parte que nos interessa.

3. Cor

Esta componente representa a média da cor de uma dada imagem/vídeo.

Tendo o tamanho da imagem, é calculado para cada pixel o valor de hue. Percorrendo todos pixéis, calcula-se a média.

Nos vídeos será uma média das frames.

4. Luminância

Esta componente representa a média da luminância de uma dada imagem/vídeo.

Se uma imagem for muito escura, irá ter uma luminância perto de 0, se for muito clara, irá ter uma luminância perto de 255.

Através do cálculo da média da cor (secção 3), conseguimos determinar a média da luminância através da seguinte fórmula:

$$L = 0.2125*R + 0.7154*G + 0.0721*B$$

R = red; G = green; B = blue

5. Número de rostos que aparecem na imagem/vídeo

Utilizando o algoritmo de deteção de objetos Haar, conseguimos determinar quantas caras aparecem numa imagem ou numa frame do vídeo.

No vídeo, para as diferentes frames, faz-se a média do número de caras.

6. Edge Distribution

Consiste no uso dos 5 filtros (figura 1) e a sua respetiva aplicação para cada imagem e, no caso dos vídeos, para algumas frames do vídeo.

Estes filtros dizem-nos qual a distribuição de edges numa imagem, isto é, se uma imagem tem mais contornos, poe exemplo, horizontais (b) ou verticais (a), sendo uma forma de caracterização da mesma.

Conseguimos perceber através dos cálculos se uma imagem é mais ou menos uniforme.

Tendo uma imagem inicial, aplicamos os filtros com as diferentes máscaras (figura 2), calculando, de acordo com a sua orientação ou direção, os contornos. Obtém-se, assim, 5 imagens diferentes, em que são contados quantos pixéis pertencem a esses contornos, isto é, quantos pixéis não são pretos, pois a imagem gerada por cada filtro vai ser toda preta, menos os contornos que estão com cores.

O resultado que é obtido irá ser a média dos pixéis pertencentes aos contornos.

No vídeo, calcula-se a média das várias frames para cada filtro.

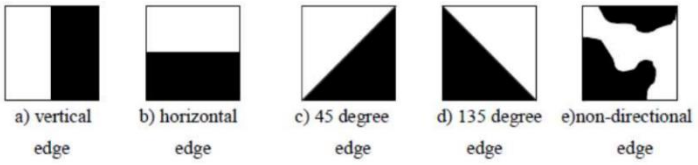


Figura 1

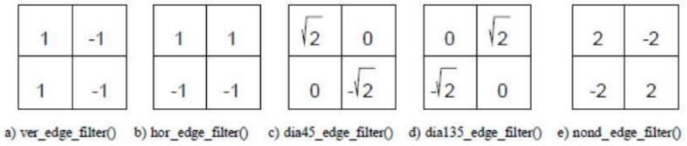


Figura 2

7. Características da Textura

Para calcular a textura usamos o filtro Gabor. Este depende de 5 parâmetros diferentes:

- Sigma, com valor 20;
- Orientation (theta), com valor 0;
- Wavelength (lambda), com valor 40;
- Aspectratio (gamma), com valor 0.5;
- Offset (phi), com valor 0.

No final é calculado 6*4 = 24 filtros Gabor. Para cada um desses filtros, calcula-se valores diferentes de média e variância.

8. Número de vezes que um objeto específico aparece

Consiste em saber se uma determinada imagem de referência aparece na imagem/vídeo.

Temos imagens de referência diferentes para as diferentes imagens e vídeos, pois seria benéfico para conseguir-se ver resultados, tendo conteúdos bastante diferentes.

Através do algoritmo ORB (rápido de ser executado), é retirado o número necessário de pontos para existir uma deteção, assim como os descritores, sendo este algoritmo bom para aplicações que exigem alterações em tempo real e em larga escala.

De seguida, consegue-se saber se determinado objeto (imagem de referência) está presente (matching) na imagem ou numa determinada frame do vídeo.

9. Ritmo

Esta componente representa todo o movimento do vídeo, sendo classificado por 3 estados possíveis: “Muda pouco”, “Muda normal”, “Muda muito”.

O cálculo que classifica esses três estados resulta de uma diferença de histogramas de frames.

Implementação da aplicação

A aplicação é carregada no modo administrador, podendo ser mudada para o modo utilizador mais tarde. Nesta interface do administrador é possível escolher entre Imagens, Vídeos, Camara e User (Utilizador).

Na interface das imagens, como nos vídeos, o utilizador pode: mudar de página, no caso de serem muitas imagens para uma página só, carregar numa imagem apenas e visualizá-la em tamanho real. Ao fazer display desse comando do lado esquerdo apresenta-se os respetivos metadados. Nos vídeos, é possível, quando selecionado, estar em modo pausa ou não, carregando apenas na tecla ‘p’.

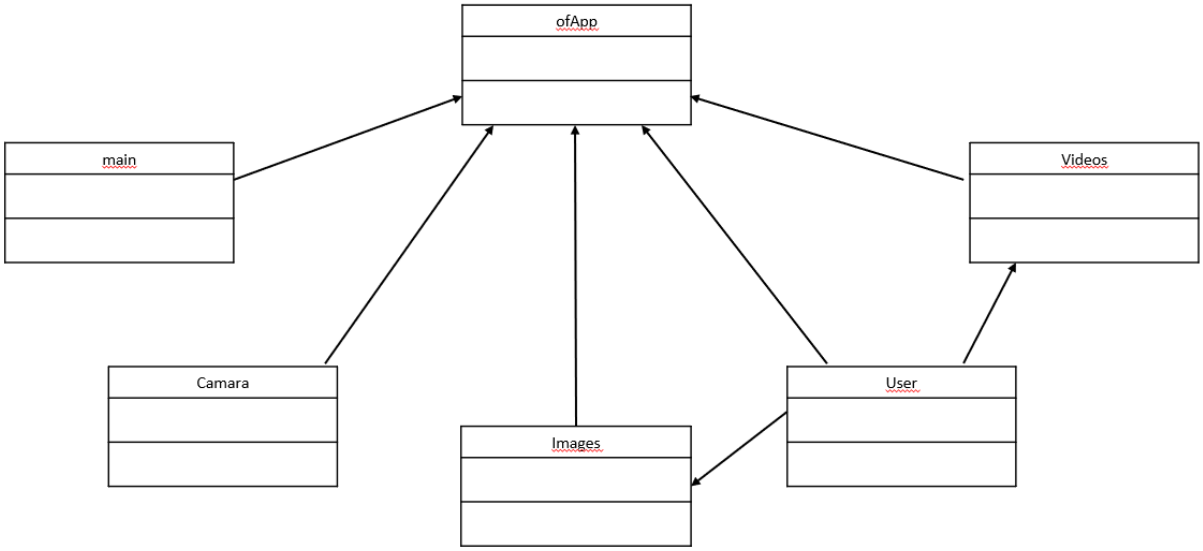
Na interface da Camara é possível visualizar o que a camara captura, assim como, o número de rostos visíveis, através dos quadrados que aparecem de volta das caras.

Na interface User são reproduzidos os vídeos e imagens correspondentes ao evento associado ao número de caras que a camara captura ou à existência de movimento ou não. O administrador pode escolher que imagens ou vídeos associar a cada evento, basta visualizar os conteúdos e modificar no nome da própria imagem ou vídeo.

Descrição das classes

Para a implementação do nosso projeto achamos que deveríamos dividir o projeto em dois caminhos diferentes: administrador e user.

A classe User é somente responsável pelo display dos vídeos e das imagens de forma alternada.



Conclusões

Para concluir, achamos que o trabalho correu bem. Contudo, durante o desenvolvimento do projeto surgiram-nos algumas dificuldades. Inicialmente, o facto de ser uma língua nova para nós tornou-se um desafio, outro impasse que tivemos durante algum tempo foi um erro que não foi fácil de solucionar.
No entanto, pensamos que o resultado final vai dar de acordo com o objetivo proposto.

Referencias

https://dl.acm.org/doi/pdf/10.1145/3025453.3025531
https://dl.acm.org/doi/10.1145/1179352.1141967
https://dl.acm.org/doi/abs/10.1145/3297156.3297184
https://dl.acm.org/doi/abs/10.1145/1504271.1504313
https://www.dcc.fc.up.pt/~mcoimbra/lectures/VC_1415/VC_1415_P8_LEH.pdf
https://docs.opencv.org/2.4/doc/user_guide/ug_features2d.html
https://dl.acm.org/doi/abs/10.1145/1073204.1073262
https://docs.opencv.org/master/d3/d63/classcv_1_1Mat.html
http://www.cplusplus.com/reference/cstdlib/abs/

Autores

Madalena Lopes, 50577
Teresa Monteiro, 52597

31/05/2020
Prof: Nuno Correia