Tag Around
Interface Gestual para Anotação de Imagens

Duarte Gonçalves[1] Rui Jesus[1,2] Filipe Grangeiro[1] Nuno Correia[1]
[1]Interactive Multimedia Group, CITI and DI/FCT, Universidade Nova de Lisboa
[2]Multimedia and Machine Learning Group, Instituto Superior de Engenharia de Lisboa
Quinta da Torre, 2829 Monte da Caparica, Portugal
dnjg@di.fct.unl.pt,rjesus@deetc.isel.ipl.pt,nmc@di.fct.unl.pt

Sumário
Este artigo descreve a interface Tag Around - um jogo de computador para anotação semi-automática de imagens baseado em interacção gestual e para ser utilizado em locais públicos. A aplicação é composta por um sistema automático de anotação de imagens, por uma interface 3D para a anotação manual e por um sistema de detecção e reconhecimento de faces. A interacção entre o utilizador e a interface é realizada através de movimentos com as mãos em frente a uma câmara. São apresentadas as várias fases de desenvolvimento da interface, incluindo a concepção da ideia, os cenários de aplicação, os testes de usabilidade efectuados e propostas para trabalho futuro.

Palavras-chave
Interface gestual, computação humana, anotação de imagens, reconhecimento facial, memórias pessoais.

1. INTRODUÇÃO

Em geral, os computadores têm sido usados para organizar todo o tipo de informação. As colecções de fotografias digitais são um exemplo de informação actualmente muito popular que pode beneficiar da capacidade organizativa de um computador pessoal. A anotação de imagens é uma operação de capital importância para uma melhor organização. Vários métodos, manuais e automáticos, têm sido usados para resolver este problema. A anotação de imagens usando palavras-chave para descrever o seu conteúdo é uma solução. ALIPR [Li06] é um sistema recente que propõe resolver o problema da anotação de imagens, usando o conteúdo (e.g., cor ou textura) para anotar automaticamente. Nos últimos anos, muitos sistemas de anotação automática têm sido propostos, contudo, o seu desempenho ainda está aquém dos sistemas manuais. As capacidades humanas são uma mais-valia neste tipo de processos, embora os seres humanos não se sintam motivados para os desempenhar [Frohlich02]. Seguindo este princípio de que é necessário motivar as pessoas a usarem as suas capacidades intelectuais nestas áreas, em [VonAhn04] foi proposto o ESP Game. Este trabalho introduziu uma nova visão para resolver o problema - a ideia de usar as capacidades humanas para completarem tarefas que os computadores ainda não conseguem desempenhar com sucesso. Várias propostas seguiram esta ideia incluindo o Manhattan Story Mashup [Tuulos07], onde vários jogadores, usando a Web, telemóveis e ecrãs públicos, tiram fotografias e anotam-nas ao mesmo tempo com o objectivo de produzir histórias em Manhattan. Outro jogo para anotação de imagens foi proposto em [Diakopoulos07]. Neste jogo é utilizado um ecrã horizontal (*tabletop*) e cada jogador utiliza um *gamepad* para jogar.

Estas abordagens trazem novas propostas ao problema da anotação de imagens, colocando o ser humano e o computador lado a lado, criando novas formas de entretenimento enquanto se organizam conteúdos multimédia. Contudo, estas abordagens limitam a experiência nos seguintes pontos:

- Utilizam dispositivos adicionais para jogar (teclado, rato, telemóveis e *gamepad*) que requerem utilizadores com algum conhecimento tecnológico;
- Não podem ser jogados em cenários diferentes por exemplo, aeroportos ou hospitais para entreter nos tempos de espera;
- Não combinam a anotação automática com a manual.

Este artigo apresenta a aplicação Tag Around, uma proposta para a anotação de imagens que consiste num jogo de computador 3D baseado em gestos, desenvolvido a pensar nas necessidades descritas anteriormente. Esta aplicação combina um sistema automático de anotação de imagens com um processo manual (através da interface do jogo), usando a detecção de movimentos manuais dos jogadores e a sua capacidade intelectual para anotar imagens. As próximas secções descrevem a aplicação e direcções para trabalho futuro.

2. TAG AROUND

A ideia de desenvolver um jogo para a anotação de imagens foi inspirada por [VonAhn04] e motivada pelas dificuldades sentidas nas experiências realizadas no âmbito do projecto Memoria [Jesus07]. Esta aplicação, como referido anteriormente, é baseada na interacção

através de gestos. Nesta secção são apresentadas as principais componentes da aplicação. Tag Around é composto por 3 blocos principais (ver figura 1): a aplicação do jogo, a interacção com o utilizador humano e um algoritmo de anotação automática de imagens. Com estas três componentes é definido um algoritmo semi-automático de anotação de imagens. Em primeiro lugar, um conjunto de imagens anotadas automaticamente (anotação automática de imagens) são apresentadas ao utilizador (interface do jogo). De seguida, por cada nova jogada (anotação) o utilizador associa um conceito a uma imagem (interacção humana) e a pontuação associada à jogada é calculada. Se um conceito tiver mais de N_a anotações, o modelo desse conceito é treinado novamente (anotação automática de imagens) com um novo conjunto de treino. Com o decorrer do jogo, mais imagens são anotadas e melhores modelos para os conceitos são estimados.

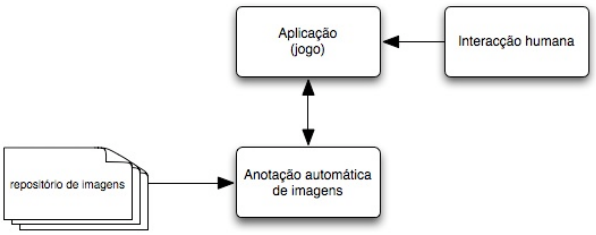

Figura 1 - Diagrama de blocos do sistema.

3. APLICAÇÃO - JOGO

A aplicação é composta por vários módulos: a interface, o motor de jogo, o módulo da detecção de movimento e o bloco de reconhecimento facial. Esta divisão justifica-se pela necessidade de adaptar a aplicação aos diferentes cenários onde pode ser utilizada (e.g., ambientes domésticos, museus, hospitais ou escolas). Por exemplo, a interface do jogo pode ser alterada, de acordo com os requisitos sociais do cenário ou a detecção do movimento pode ser modificada consoante as condições do local. As secções seguintes descrevem os diferentes módulos da aplicação.

3.1 Interface

A interface do jogo foi concebida usando o OGRE (Object-oriented Graphics Rendering Engine), um motor de de gráficos 3D orientado para objectos e multi-plataforma.

O *layout* inicial do jogo é constituído por um menu em que o utilizador (usando gestos) poderá escolher entre 2 opções, *Play Game* e *Highscores* (ver figura 2). Uma vez no modo de jogo, o utilizador entra no *layout* de reconhecimento facial (ver secção 3.4) para efectuar o *login*, e de seguida inicia o jogo (ver figura 3).

A interface do jogo representada na figura 3 é constituída pelos seguintes elementos dispostos no ecrã: a imagem do utilizador com as zonas sensíveis, um conjunto de anotações colocadas numa plataforma giratória, um conjunto de imagens na parte inferior do ecrã, uma barra de energia (que permite ao utilizador saber quando está próximo do fim do jogo), a pontuação (que varia consoante o jogador faz boas ou más anotações) e uma lista de anotações já efectuadas pelo utilizador na imagem seleccionada.

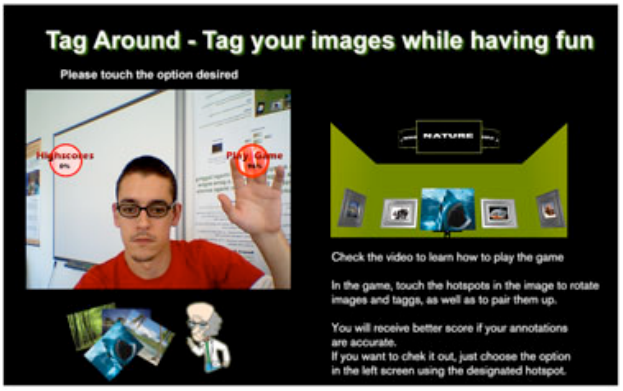

Figura 2 - *Layout* inicial do jogo.

Quando o jogo termina (a barra de energia desaparece), a pontuação, o número de anotações assim como a confiança que o sistema tem no utilizador é mostrada, e o perfil do jogador é guardado, para futuras utilizações por parte do mesmo utilizador.

Figura 3 - Interface do jogo.

3.2 Motor de jogo

A dinâmica do jogo é a seguinte: quando o jogo começa é activado um temporizador e um conjunto de imagens seleccionadas aleatoriamente é apresentado ao utilizador; a seguir o utilizador tem de anotar a maior quantidade possível de imagens com conceitos; quando termina o tempo estipulado para cada nível, novas imagens são apresentadas; o jogo acaba quando a energia chega ao fim.

No início do jogo, o utilizador terá três minutos para anotar 5 imagens com 8 conceitos diferentes e, à medida que os níveis avançam, menos tempo terá para anotar as suas imagens. A barra de energia no topo do ecrã está relacionada com o tempo que resta ao utilizador para anotar as imagens. Quanto melhores anotações forem efectuadas mais energia será atribuída. Por outro lado, o tempo e as más anotações irão fazer com que a barra de energia diminua e consequentemente o jogo termine. Em termos de anotações, quanto melhores as anotações melhor pontuação e confiança o utilizador irá ter, o que lhe permitirá obter melhores pontuações em anotações futuras.

O jogador recebe uma pontuação mais alta se fizer boas anotações ("boa jogada"). Uma anotação é pontuada com base em três valores obtidos por: (1) o algoritmo automático de anotação de imagem (ver secção 5); (2) a jogada do utilizador e a confiança que o sistema tem no mesmo; (3) e o *feedback* do grupo de utilizadores que já anotou essa imagem previamente.

Para um jogador que efectue a primeira anotação numa imagem, a pontuação depende exclusivamente do algoritmo automático de anotação de imagem. É-lhe atribuída uma pontuação e a confiança que o sistema tem nele está de acordo com a percentagem que o algoritmo deu a essa anotação nessa imagem particular.

Quando um grupo de utilizadores efectua várias anotações em imagens, a pontuação de um utilizador é influenciada maioritariamente pelo feedback dos outros utilizadores, tornando-se assim um sistema de anotação social e manual, que irá posteriormente cruzar resultados com o sistema automático, de forma a treinar e aperfeiçoar esse mesmo sistema de acordo com os resultados obtidos.

3.3 Detecção de movimento

Esta interface é baseada na interacção usando gestos (movimentos manuais simples). Para este tipo de interacção, foi usado o OpenCV (Intel Open Source Computer Vision Library), e foram testados vários algoritmos de detecção de movimento e processamento de imagem para detectar em que zonas da imagem do utilizador (*hotspots*) existe movimento. Foram experimentados algoritmos baseados em *optical flow* e em *motion detection*. Pelos testes efectuados optou-se pela técnica de *motion detection* dado que em termos de interacção é mais simples para o utilizador.

3.4 Reconhecimento facial

Neste módulo efectuam-se os algoritmos de processamento de imagem para detectar e reconhecer a face do jogador. Este módulo foi dividido em três tarefas: detecção, normalização e reconhecimento de faces.

Em primeiro lugar, é preciso detectar a presença de uma face na imagem capturada pela câmara. O método utilizado é baseado no sistema descrito em [Viola04] complementado por um método de detecção de pele para confirmar a presença de faces.

Para normalizar as imagens das faces utilizou-se a equalização de histogramas para resolver os problemas de iluminação e a detecção de olhos para regularizar a posição dos mesmos na imagem da face detectada

Finalmente, para o reconhecimento facial utilizou-se uma técnica para representação das faces e outra para classificação das identidade das faces através do método descrito em [Muller01]. Para ajudar este processo de reconhecimento procedeu-se à estimativa da pose da face através do método [Viola04].

4. INTERACÇAO HUMANA

A aplicação proposta baseia-se em interacção gestual. O utilizador encontra-se de frente para uma webcam e através de gestos, roda as anotações e as imagens, com o intuito de as ligar, para formar pares anotação-imagem. Para registar novos jogadores é utilizada uma interface de reconhecimento de faces.

4.1 Interacção gestual

Para jogar o utilizador tem de efectuar movimentos com as mãos em zonas sensíveis (ver figura 3). Estes movimentos são capturados por uma câmara e analisados através de métodos de detecção de movimento. Esta forma de navegar interactivamente é feita em algumas zonas no ecrã (denominadas de *hotspots*), que numa primeira fase lhe darão acesso tanto ao jogo como a um *layout* de *highscores*. Numa segunda fase o jogador usa as zonas sensíveis para rodar anotações e imagens, a fim de formar pares anotação-imagem, e assim conseguir alcançar uma pontuação máxima no jogo.

4.2 Interface de reconhecimento facial

Neste sistema é difícil manter e actualizar informação sobre cada utilizador uma vez que a interacção não se faz através das técnicas mais habituais como o rato ou o teclado. A solução utilizada para registar utilizadores e efectuar *logins* é baseada no reconhecimento facial do jogador.

A utilização desta interface consiste na colocação da face numa área limitada por um quadrado durante 10 segundos para que o sistema proceda ao seu reconhecimento (ver secção 3.4). Durante esse tempo, é mostrado o estado de evolução do processo sob a forma de percentagem. Uma ilustração da interface de *login* pode ser vista na figura 4.

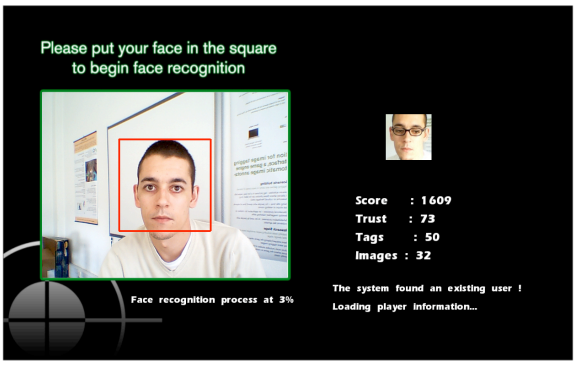

Figura 4 - Interface perceptual para registo de novos utilizadores.

A aplicação de um método de reconhecimento de faces neste sistema depara com os seguintes problemas: dificuldade no reconhecimento da identidade da pessoa correcta devido à quantidade e variabilidade reduzida de fotos de cada pessoa nos primeiros *logins* e reconhecimento indevido das pessoas que, por exemplo, poderão estar a assistir ao jogo.

Estas dificuldades foram resolvidas da seguinte forma: o reconhecimento do utilizador é feito durante 10 segundos (300 imagens), sendo guardadas novas fotos do utilizador a cada login. É também limitado o reconhecimento de faces a uma área quadrada indicada.

5. ANOTAÇÃO AUTOMÁTICA DE IMAGEM

O algoritmo automático de anotação de imagens calcula a probabilidade de um conceito (*tag)* dada uma imagem. Para o conjunto definido previamente de conceitos apresentados ao utilizador é calculado um modelo probabilístico. Estes modelos são treinados utilizando características visuais extraídas automaticamente de imagens de treino. Assim, novas imagens são automaticamente classificadas de acordo com estes modelos. No início, toda a base de dados é classificada com todos os modelos treinados. Mais detalhes sobre esta aproximação são descritos em [Jesus07].

6. DESIGN DA INTERFACE

O projecto e a implementação da interface do jogo para anotação de imagens foi feito de forma iterativa. No início foram definidas as ideias principais e vários cenários de aplicação. De seguida foram realizados vários testes com protótipos em papel e finalmente foram realizados testes de usabilidade. O protótipo computacional foi refinado em cada uma das fases de forma iterativa.

6.1 Protótipo em papel

Como referido anteriormente, uma das questões fulcrais da aplicação foi o uso de uma câmara e a sua interacção com os movimentos do utilizador com o objectivo de anotar imagens. O protótipo em papel incluiu uma série de tarefas apresentadas aos utilizadores, que tinham uma noção prévia do conceito da aplicação mas desconheciam o módulo da interacção gestual. Foram elaborados 5 testes de usabilidade com estudantes da área, todos com experiência no uso dos computadores. Os testes foram divididos em duas fases. Numa primeira fase os utilizadores tinham de anotar imagens sem as restrições a nível de tempo e pontuação sendo que posteriormente os utilizadores teriam a informação da sua pontuação e do tempo disponível.

Após a análise dos testes referidos, a interface foi refinada e um protótipo computacional foi desenvolvido com base nas seguintes observações:

- Inicialmente alguns utilizadores tiveram dificuldades com a interacção com a câmara. Isto aconteceu quando tentavam rodar as imagens e os conceitos e simultaneamente controlar a pontuação e o tempo de jogo. Contudo, depois de algumas experiências ultrapassaram estas dificuldades e começaram a divertir-se;
- Alguns utilizadores perguntaram se era possível adicionar elementos visuais, para além da pontuação, para indicar se a anotação tinha sido boa ou má.

Como consequência, foi adicionada à interface uma pequena animação e um comentário sempre que é feita a anotação de forma a indicar ao utilizador a qualidade da anotação.

6.2 Testes de usabilidade

Após a elaboração da interface, estão a ser realizados uma série de testes de usabilidade em diferentes cenários: numa escola, num local público (aeroporto) e no campus da universidade. Estes testes têm com objectivo perceber se a interface, módulos de interacção, motor de jogo (pontuação, confiança nos utilizadores) e módulo de reconhecimento facial estão de acordo com os padrões dos utilizadores, e se a própria interface é apelativa para os mesmos, de forma a proporcionar uma forma divertida e simples de anotar imagens.

7. CONCLUSÕES E TRABALHO FUTURO

Este artigo apresentou uma aplicação para anotação semi-automática de imagens baseada em interacção gestual e para ser utilizada em locais públicos. No futuro irão ser realizados mais testes de usabilidade em diferentes cenários e a aplicação irá ser aperfeiçoada, de acordo com os requisitos e resultados dos testes para cada cenário.

8. REFERÊNCIAS

[Diakopoulos07] Diakopoulos, N., Chiu, P., PhotoPlay: A Collocated Collaborative Photo Tagging Game on a Horizontal Display. *In Proceedings addendum of User Interface Software Technology (UIST)*, (2007).

[Frohlich02] Frohlich, D., Kuchinsky, A., Pering, C., Don, A., and Ariss, S., "Requirements for Photoware," Proc. of the ACM conference on Computer supported cooperative work, (2002), 166-175.

[Jesus07] Jesus, R., Dias, R., Frias, R., Abrantes, A., Correia, N.,. Sharing Personal Experiences while Navigating in Physical Spaces. *In 5th Workshop on Multimedia Information Retrieval, proceedings of the 30th international ACM Information Retrieval Conf (SIGIR)*, (2007).

[Li06] Li, J., Wang, J., Real-time computerized annotation of pictures. In *ACM Intl. Conf. on Multimedia*, (2006), 911-920.

[Muller01] Muller, K.-R., Mika, S., Ratsch, G., Tsuda, K., Scholkopf, B.,. An Introduction to Kernel-Based Learning Algorithms Vol.12 (2), (2001), p. 181-201.

[Tuulos07] Tuulos, V., Scheible, J., Nyholm, H. , Combining Web, Mobile Phones and Public Displays in Large-Scale: Manhattan Story Mashup. In *Proceedings of the Fifth International Conference on Pervasive Computing*, (2007).

[Viola04] Viola P., Jones M., Robust Real-Time Face Detection. *International Journal of Computer Vision*, Vol. 57(2), (2004), p. 137–154.

[VonAhn04] von Ahn, L., Dabbish, L. , Labeling images with a computer game. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems CHI '04*, (2004), 319-326.