Extracção de dados sobre o formato dos lábios nos sons da Língua Portuguesa

Alexandra Grilo, Pedro Duarte, António Teixeira, Armando Pinho

***Resumo* - Este artigo aborda o problema da extracção automática da forma dos lábios correspondente a sons característicos da Língua Portuguesa. Com essa finalidade, foi desenvolvido um sistema de aquisição e análise de imagem baseado em algoritmos genéticos, os quais, conjuntamente com outras técnicas mais tradicionais, proporcionam a segmentação da zona da boca. Com base nessa segmentação são calculados alguns parâmetros relevantes para o problema abordado.**

***Abstract* – This paper addresses the problem of automatic extraction of lips format parameters for Portuguese sounds. For that purpose, a system for image acquisition and analysis was developed. The system, based in genetic algorithms and more traditional techniques, calculates several relevant parameters for the lips, using the segmented mouth region.**

I. INTRODUÇÃO

Assiste-se cada vez mais ao desenvolvimento de sistemas de síntese utilizando além da voz uma face sintética. As duas modalidades associadas permitem uma maior imunidade ao ruído e maior inteligibilidade. Com o desenvolvimento de sistemas de telecomunicações como o vídeo telefone, pode encarar-se a possibilidade de dotar as máquinas com este tipo de interface, usando "visual speech" [5].

Para isso, é necessário adquirir conhecimento sobre a configuração dos lábios para os vários sons da língua portuguesa. Numa fase posterior, estes conhecimentos podem ser utilizados na construção de um modelo tridimensional dos lábios e finalmente da cara [7].

O objectivo deste trabalho é a implementação de um sistema para a obtenção, de forma automática, de dados sobre a configuração dos lábios durante a locução dos diversos fonemas da língua portuguesa, com especial incidência nos sons vocálicos [9]. Em termos técnicos este projecto envolveu a aquisição de vídeo e posterior aplicação de algoritmos de processamento de imagem para a extracção de contornos [6][10].

II. AQUISIÇÃO

No início deste trabalho, foi proposto implementar um sistema de aquisição sincronizada de áudio e vídeo onde eram utilizadas duas câmaras para obter uma imagem lateral conjuntamente com a frontal. No entanto, tal não se veio a verificar por tal sistema ser demasiado complexo e dispendioso.

Outra das soluções possíveis seria o uso de duas câmaras em simultâneo [1]. No entanto esta solução tinha o inconveniente de não se conseguir o sincronismo automático, isto é, as imagens adquiridas por ambas as câmaras necessitavam de um sinal para se efectuar o sincronismo.

A solução por nós utilizada foi a mais simples e menos dispendiosa, pois utilizava apenas uma câmara para adquirir a imagem frontal. A câmara utilizada foi uma Quick Cam da Connectix que adquiria imagens e vídeos com uma resolução de 320x240 e uma *frame rate* da ordem das 10 *frames* por segundo.

III. SEGMENTAÇÃO

Geralmente, o primeiro passo a dar na análise de uma imagem passa pela segmentação desta. Tal tarefa significa dividir a imagem nos seus diferentes conteúdos. No nosso caso específico, o conteúdo da imagem a segmentar reside na zona interna da boca. Neste trabalho apresentamos dois métodos distintos embora com o mesmo objectivo: retirar de uma imagem facial a informação referente à zona interna da boca quando é pronunciado um som pelo indivíduo filmado.

O primeiro método aqui apresentado, o qual se baseia em algoritmos genéticos, aborda o tema da detecção de contornos de uma forma bastante diferente das dos métodos convencionais, tais como os filtros de Sobel, Prewitt e Frei-Chen [6]. Ao contrário destes métodos, o algoritmo genético tenta detectar os contornos sem processar por completo a imagem, baseando-se na forma como os cromossomas interagem entre si.

A. Segmentação baseada em algoritmos genéticos

No processamento digital de imagem, características tais como contornos, linhas e curvas podem ser detectados utilizando para tal alguns operadores matemáticos entre os quais gradientes, passagens por zero e filtros. A detecção dessas características pode facilitar bastante a compreensão da imagem em questão. No entanto, estas abordagens convencionais exigem o processamento de toda a imagem, processo esse que se pode tornar moroso.

O algoritmo aqui proposto utiliza agentes autónomos que se podem auto-reproduzir, mover e morrer durante a interacção com a imagem digital [8]. Desta forma pretende-se obter um método descentralizado baseado em

comportamentos bem definidos, não havendo por isso necessidade de proceder à análise de toda a imagem.

1. Agentes autónomos

Os agentes autónomos operam num ambiente rectangular correspondente à imagem digitalizada. Na imagem digitalizada, cada grelha de oito pixels conectados representa uma localização possível para o agente habitar, quer definitivamente, quer temporariamente (Fig. 1).

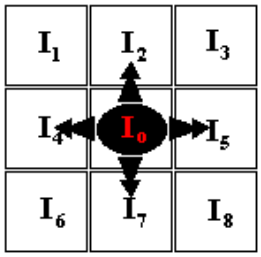

Figura 1 - Cada pixel no ambiente 2D é conectado a oito pixels vizinhos (*k*=1). Um agente autónomo pode-se auto-reproduzir ou mover em cada um dos oito pixels vizinhos.

A região vizinha de um agente num pixel *p* é uma região circular centrada no pixel *p* com raio *k*. Os pixels que se situam nesta região são chamados pixels vizinhos ao agente.

Um agente autónomo escolhe o comportamento a adoptar após o estudo da concentração de certos níveis de cinzento na sua região vizinha. Se a concentração estiver dentro de uma determinada gama, o agente activará o seu mecanismo de auto-reprodução. Tal concentração é designada por estímulo local.

O estímulo local que selecciona e activa o comportamento do agente é calculado através da distribuição da densidade de todos os pixels da região vizinha cujo níveis de cinzento se localizem perto da intensidade do pixel onde o agente se localiza. Mais especificamente, a distribuição de densidade é definida por

$$D^{K}_{I(i,j)} = \sum_{s=-k}^{k}\sum_{t=-k}^{k}\left\{\frac{1}{\|I(i+s,\, j+t) - I(i,\, j)\|} < \delta\right\}$$

onde *k* representa o raio da região de vizinhança centrada no ponto de agente (i,j), s e t indíces de um pixel pertencente à vizinhança, I(i,j) o nível de cinzento em (i,j) e δ um limiar pré-definido [8].

Durante a avaliação da região vizinha, cada agente apresentará um determinado comportamento. Esse comportamento é accionado pelo estímulo externo presente no ambiente que o rodeia. Quando um agente detecta um pixel de contorno, é colocado um marcador nesse pixel. Sendo λ=[u,v] uma gama aceitável, onde u ≤ v, o agente coloca um marcador no pixel *p* se a avaliação da distribuição da densidade em *p* cair dentro da gama especificada por λ.

De acordo com a distribuição da densidade de um dado agente numa dada região o agente adoptará um dos seguintes comportamentos:

Difusão: Sendo λ =[u,v] um intervalo aceitável para a contagem de pixels usando (1), o agente mover-se-á para as localizações adjacentes cada vez que a distribuição de densidade cair fora do intervalo λ, i.e., $D^{k}_{I(i,j)} \notin \lambda$. A direcção da difusão será escolhida aleatoriamente na primeira vez que o agente se deslocar. A partir do momento em que uma direcção é escolhida para um agente, esse agente deslocar-se-á sempre nessa direcção (Figura 2).

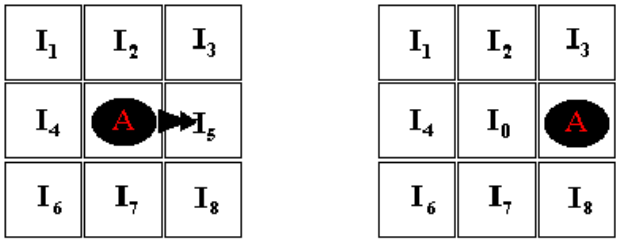

Figura 2 - Caso $D^{k}_{I(i,j)} \notin \lambda$ então o agente deslocar-se-á numa direcção determinada no primeiro movimento.

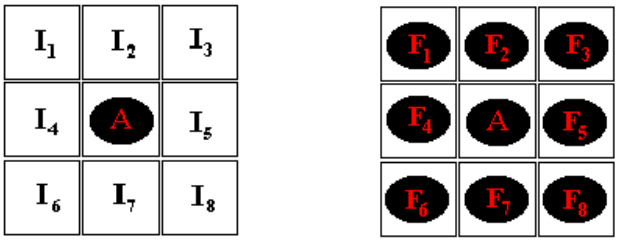

Figura 3 - Caso $D^{k}_{I(i,j)} \in \lambda$ então o agente produzirá agentes filho na sua zona de vizinhança.

Auto-reprodução: Se um agente detecta um pixel de contorno *p* ($D^{k}_{I(i,j)} \in \lambda$), produzirá um número finito de agentes filho dentro da zona de vizinhança de raio *k*. Neste trabalho foi por nós considerado que sempre que um agente detecta um pixel de contorno seriam produzidos por esse agente oito agentes filho (Fig. 3).

Morte: Quando um agente chega ao fim do seu tempo de vida, não pode proceder a mais nenhuma operação e é apagado do ambiente. Assim a detecção do contorno continuará a ser efectuada pelos seus descendentes, que por sua vez irão dar lugar a uma nova prole caso detectem o contorno. Caso o agente se desloque para uma zona onde não haja contorno, esse agente morrerá sem dar origem a uma nova geração.

2. Algoritmo

De seguida é apresentado o algoritmo completo utilizado para a detecção dos contornos dos lábios.

```
Distribuir aleatoriamente n agentes pela imagem,
inicializando o seu tempo de vida {φ_n^0}.
While ((NumAgentesVivos ≠ 0) and
      ((NumOperaçõesEfectuadas ≤ VAL_MAX))
   For  i = 0  to NumAgentesVivos
      If  ( D^k_I(i,j)(φ_i ) ∈ λ ) then
         Gerar 8 agentes filhos na vizinhança.
         NumAgentesVivos=NumAgentesFilho + 8.
      Else
         If ( Primeiro movimento do agente )
               Calcular  direcção do movimento
         End If
         Mover agente φ_i
      End If
      If ( IdadeDoAgente = MAX_IDADE ) then
         Matar agente φ_i
         NumAgentesViVos--
      Else
         IdadeDoAgente++
      End If
   Next i
End While
```

3. Resultados obtidos

Utilizando o algoritmo atrás descrito, começámos por aplicá-lo numa imagem de teste bastante simples com vista a analisar quais as suas verdadeiras capacidades na detecção de contornos, por mais simples que estes fossem. A evolução do algoritmo ao longo das diferentes operações está descrita nas imagens apresentadas nas Figuras 4 e 5.

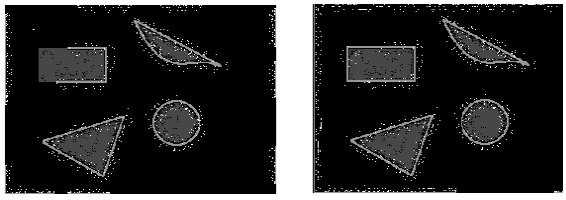

Figura 4 - Duas etapas na detecção dos contornos utilizando o algoritmo genético.

Os resultados foram obtidos utilizando 500 agentes iniciais, δ = 20 e λ = [1,7]. As linhas mais claras representam os agentes que se encontram sobre pontos de contorno, enquanto que os pontos claros dispersos pela imagem representam agentes que não estão mortos, mas que ainda não encontraram pontos de contorno.

Como se pode ver na Figura 4, verifica-se que os contornos presentes na imagem de teste são correctamente detectados (n=70). Os agentes, ao encontrarem pontos de contorno, marcam esses pontos e começam a formar colónias nessa zona, colónias essas que evoluem em torno das zonas de contorno. Os agentes que não encontram zonas óptimas de estabelecimento acabam por morrer, desaparecendo da imagem de teste.

Embora os resultados obtidos sejam satisfatórios, é importante salientar que a imagem de teste utilizada é bastante contrastada, facilitando a detecção dos contornos.

Como o objectivo deste projecto é detectar os contornos internos dos lábios, procedemos de seguida à detecção dos contornos em imagens faciais. Os resultados obtidos estão

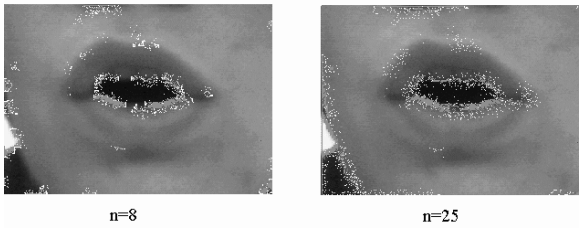

Figura 5 – Aplicação do algoritmo genético às imagens faciais.

expostos na Figura 5.

B. Segmentação baseada na Análise do Histograma

Ao contrário do algoritmo apresentado atrás, o algoritmo baseado no histograma vai isolar determinadas características da imagem, com vista a detectar os parâmetros dos lábios. Uma diferença importante entre os dois algoritmos reside no facto do algoritmo do histograma necessitar de processar toda a imagem.

1. Cálculo do histograma de uma imagem

O histograma de uma imagem digital a preto e branco é uma função discreta

$$p(r_k) = \frac{n_k}{n}$$

onde r_k representa o nível de cinzento k, n_k representa o número de pixels da imagem com esse nível de cinzento e n o número total de pixels na imagem.

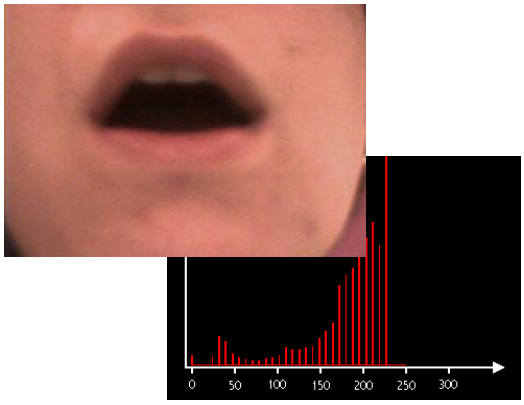

Figura 6 - Imagem típica utilizada e respectivo histograma.

Desta forma, $p(r_k)$ dá-nos uma estimativa da probabilidade de ocorrência do nível de cinzento r_k. Um

gráfico desta função para todos os valores de k fornece uma descrição global do conteúdo da imagem. Dependendo das características da imagem, o histograma pode apresentar várias formas.

Calculando o histograma para o nosso tipo de imagem, obtivemos os resultados que se apresentam na Figura 6.

Pela observação do histograma, constata-se que há um pico no intervalo de 0 a 70 que nos representa os níveis de cinzento mais escuros. Na imagem, os níveis mais escuros representam o interior da boca. Utilizando estes dados, tentámos isolar a zona interna da boca, usando para este fim o algoritmo a seguir apresentado.

2. Algoritmo utilizado

Utilizando os conhecimentos adquiridos pela análise do histograma, desenvolvemos um algoritmo capaz de segmentar o interior da boca, rejeitando tudo o resto.

De seguida é apresentado o algoritmo completo utilizado para a detecção dos contornos dos lábios :

```
For  x = 0 to DimX
 For y = 0 to DimY
   If ValMinCinzento<Imagem(x,y)< ValMaxCinzento
   Then Marcar ponto como PontoInternoDaBoca
 Next y
Next x
```

3. Resultados obtidos

Aplicando o algoritmo acima descrito sobre imagens faciais, os resultados obtidos foram semelhantes aos apresentados na Figura 7.

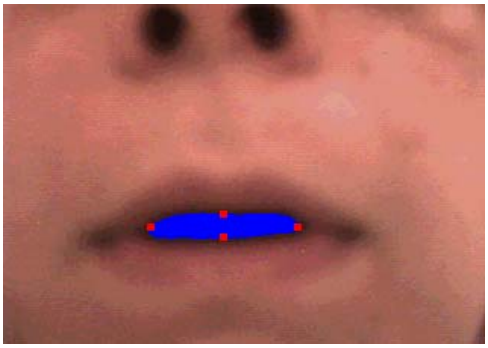

Figura 7 - Aplicação do algoritmo do histograma.

De modo a só se detectar o interior da boca, foi necessário isolar parte da imagem, visto que se não houvesse isolamento, também as fossas nasais seriam detectadas.

IV. CÁLCULO DOS PARÂMETROS DOS LÁBIOS

A. O problema da escala da imagem

Um dos primeiros problemas que se nos depara quando tentamos calcular os parâmetros dos lábios prende-se com o facto de durante a aquisição das imagens o indivíduo filmado poder estar mais afastado ou mais perto da câmara. Para tal é necessário ter um método que nos permita calcular a escala das distâncias.

O método por nós utilizado consiste em medir a largura da boca do indivíduo em repouso (em pixels) e, tendo como base a distância real em centímetros, calcular um factor que permita saber a quantos pixels corresponde um milímetro. Este factor é chamado de factor de escala e será mencionado posteriormente aquando da análise da aplicação ***Magic Lips***.

A largura da boca é calculada utilizando para tal o método do histograma visto anteriormente. No entanto, o limite de cinzento máximo aqui utilizado passa de 70 para 100. Este cálculo deve ser efectuado antes de iniciar qualquer processamento.

B. Contorno do algoritmo genético

Após a informação fornecida pelo algoritmo genético, torna-se necessário processar a imagem por forma a obter desta a informação acerca da altura, largura e área da zona interna da boca. A informação processada pelo algoritmo genético é passada através de uma matriz do tamanho da imagem processada, matriz esta que contém os pontos de contorno.

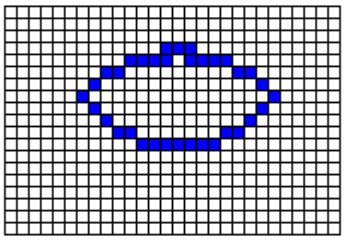

Figura 8 - Forma matricial como o algoritmo genético representa o contorno.

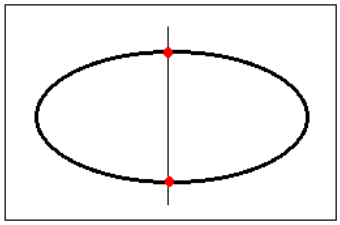

Figura 9 - Cálculo da altura da boca através de pontos extremos do contorno.

Para calcular a altura da boca é necessário calcular as coordenadas dos pontos de contorno que se afastam o mais possível na vertical. Considerando que a boca se encontra centrada na imagem, e considerando o ponto central da imagem, ao deslocar esse ponto na vertical (tanto para cima como para baixo), esse ponto irá encontrar os pontos extremos do contorno na vertical. A altura da boca será o módulo da diferença entre os pontos encontrados.

V. A APLICAÇÃO MAGIC LIPS

Foi desenvolvida, usando Visual Basic, uma aplicação integrando as facilidades de manipulação de filmes e os algoritmos de extracção de parâmetros dos lábios.

Na Figura 10 apresenta-se a interface da aplicação, denominada ***Magic Lips***.

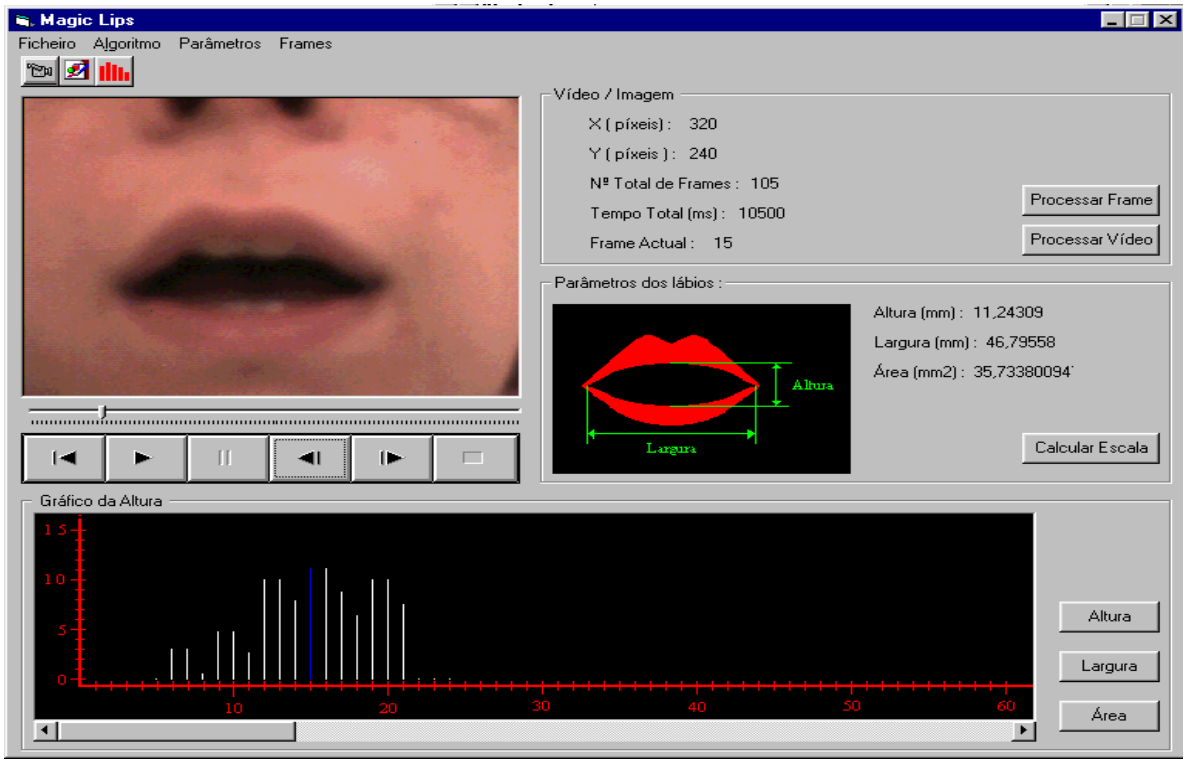

Figura 10 - Interface da aplicação ***Magic Lips***.

A interface da aplicação é composta por diferentes áreas, cada qual com uma função específica:

- menus para: operações com ficheiros, escolha do algoritmo, ajuste dos parâmetros dos algoritmos e selecção da parte de um vídeo que o utilizador pretende analisar. Existem também botões para acesso rápido a algumas das opções anteriores.
- zona de visualização das imagens. Esta zona permite manipular as imagens. Para tal existe uma zona de botões destinada a efectuar operações sobre o vídeo.
- zona de visualização dos parâmetros de vídeo/imagem. Permite o processamento de uma *frame* ou de um vídeo.
- A zona de parâmetros dos lábios. Após o processamento de cada *frame*, os parâmetros são disponibilizados nesta zona.
- A zona de visualização dos gráficos. Permite visualizar a variação dos parâmetros dos lábios ao longo do tempo. Três botões permitem comutar entre os gráficos da altura, da largura e da área.

VI. APLICAÇÃO AO ESTUDO DE VOGAIS

A. *Estudo de vogais isoladas*

O objectivo deste trabalho reside em extrair dados sobre o formato dos lábios nos sons da língua portuguesa, incidindo nas vogais. Analisámos vídeos onde um sujeito produzia os sons referentes a cinco vogais da língua portuguesa ([a], [E], [i], [□], [u])[1]. A vogal [E] corresponde ao som da palavra p**é** e a vogal [□] ao som utilizado na palavra p**ó**. As outras três aparecem em p**á,** p**i,** p**u**lo.

Os valores obtidos estão expostos nas tabelas seguintes:

Método do histograma

Vogal	Altura (mm)	Largura (mm)	Área (mm^2)
[a]	7.98	38.76	286.67
[E]	4.49	34.23	150.64
[i]	2.46	30.65	87.25
[□]	7.65	34.87	240.56
[u]	3.73	21.09	62.59

Algoritmo genético

Vogal	Altura (mm)	Largura (mm)	Área (mm^2)
[a]	10.54	49.02	405.92
[E]	7.02	49.67	273.66
[i]	4.65	53.92	196.94
[□]	10.77	55.28	467.77
[u]	1.87	57.94	85.04

Os vídeos utilizados foram comuns aos dois métodos, no entanto existem discrepâncias entre alguns valores obtidos, nomeadamente nas larguras (com óbvias

repercussões nos valores das áreas). Esta ocorrência deu-se devido às condições de luz verificadas durante a filmagem dos vídeos. A câmara utilizada era muito sensível à quantidade de luz, sendo muito difícil obter imagens com boa qualidade. Desta forma resolvemos utilizar apenas os valores obtidos com o método do histograma, uma vez que este método veio a revelar-se menos sensível à qualidade das imagens.

Os resultados obtidos estão qualitativamente de acordo com a descrição, em termos do posicionamento dos lábios, que é geralmente feita das vogais do Português [2]. As vogais [a] e [□] são as que apresentam maior abertura (altura) e maiores áreas de abertura dos lábios. A vogal [u] devido à configuração arredondada dos lábios e consequente protrusão (movimento para a frente) apresenta uma área reduzida e também, pelo menos segundo o método do histograma, largura da abertura dos lábios inferior às outras vogais.

B. Estudo de vogais integradas em palavras

O estudo efectuado na alínea anterior visava as vogais isoladas, não havendo desta forma ligação entre estas e outros sons da língua portuguesa. Neste trabalho procurou-se identificar os parâmetros dessas mesmas vogais num contexto ligeiramente diferente e verificar se existem diferenças significativas. Para tal utilizámos a palavra *pato*, palavra esta que tem a uma expressão clara das vogais [a] e [u]. A figura seguinte mostra as imagens retiradas do vídeo e referem-se ao momento em que as vogais são pronunciadas. Para tal foi utilizado o algoritmo do histograma, visto este detectar com exactidão os pontos extremos das áreas em questão.

Figura 4 - Imagens relativas às vogais analisadas no vídeo da palavra *pato*, à esquerda o [a], à direita o [u].

Os resultados obtidos, semelhantes aos obtidos para estas duas vogais quando proferidas isoladamente, são apresentados na tabela seguinte.

Vogal	Altura	Largura	Área
[a]	9.721	39.40	369.22
[u]	1.790	23.79	56.41

VII. CONCLUSÕES

A segmentação baseada em algoritmos genéticos revelou adaptar-se bem às características das imagens em questão. É preciso, no entanto, mais trabalho para resolver o problema da determinação da largura de abertura dos lábios. As grandes vantagens deste algoritmo são, principalmente, não ser necessário processar toda a imagem e poder utilizar-se informação das *frames* anteriores. Estas características são muito interessantes para a aplicação pretendida. Refira-se que este algoritmo também permite a obtenção de informação do contorno exterior dos lábios e mesmo do rosto, informação essa necessária para sistemas de *visual speech*.

O algoritmo do histograma, embora conceptualmente bastante mais simples do que o algoritmo genético, revelou-se bastante rápido e eficiente, tendo produzido resultados muito bons.

Pudemos constatar que a extracção de parâmetros descrevendo o formato dos lábios não é uma tarefa fácil. No entanto, sendo uma área relativamente recente, foi gratificante concluir este trabalho com alguns dos objectivos atingidos. Por outro lado, embora o material utilizado fosse satisfatório, acreditamos que a utilização de imagens de melhor qualidade deverá contribuir para uma melhoria dos resultados.

Como trabalho futuro, identificamos as seguintes questões que poderão ser abordadas ou aprofundadas: resolução dos problemas detectados no algoritmo genético; utilização do algoritmo genético na obtenção de outros parâmetros da face; efectuar um estudo exaustivo de todas as vogais e consoantes existentes na língua portuguesa e utilização dos dados no desenvolvimento de modelo tridimensional dos lábios e/ou da face.

REFERÊNCIAS

[1] A. Andrade e M. C. Viana, "Fonética", *em Introdução à Linguística Geral e Portuguesa*, I. H. Faria et al., Caminho, 1996.

[2] J. M. Barbosa, *Introdução ao Estudo da Fonologia e Morfologia do Português*, Almedina, 1994.

[3] D. Beasley, D. R. Bull e R. R. Martin, "An Overview of Genetic Algorithms" , *University Computing*, 1993.

[4] F. Lavaghetto, S.Lepsoy, C. Braccini e S. Curinga , "Lip Motion Modeling and Speech Driven Estimation", Proc. ICASSP, 1997.

[5] B. Le Goff, T. Guiard Marigny e C. Benoît, "Analysis-Synthesis and Intelligibility of a Talking Face", *Em Progress in Speech Synthesis*, J. van Santen, et. al. (Editores),, Springer Verlag, 1997.

[6] R. C. Gonzalez, R. E. Woods, *Digital Image Processing*, Addison-Wesley Publishing Company, 1992.

[7] T. Guiard-Marigny, Ali Adjoudani e C. Benoît , "3D Models of the Lips and Jaws for Visual Synthesis", *Em Progress in Speech Synthesis*, J. P. H. Van Santen, et. Al (Editors), 1997.

[8] J. Liu, Y. Y. Tang e Y. C. Cao, "An Evolutionary Autonomous Agents Approach to Image Feature Extraction" , *IEEE Trans. on Evolutionary Computation*, Vol.1, Nº 2, Julho 1997.

[9] M. R. D. Martins, *Ouvir Falar – Introdução à Fonética do Português*, Caminho, 1988.

[10] W. K. Pratt, *Digital image processing*, Wiley-Interscience, 1991.

[11] D. Whitley, *A Genetic Algorithm Tutorial*, Colorado State University, 1993.