

Base de dados voltada à análise das variações faciais causadas pelo envelhecimento

Roberto Sales Caldeira, Maurício Pamplona Segundo
Intelligent Vision Research Lab
Universidade Federal da Bahia
Salvador, Bahia

E-mail: robertosalesc@dcc.ufba.br, mauricio@dcc.ufba.br

Resumo—O objetivo deste trabalho é desenvolver uma base de imagens voltada ao estudo do processo de envelhecimento facial, tendo em vista o avanço nos métodos de reconhecimento facial de indivíduos sob variação de idade. Para tal, foi realizada uma coleta de imagens de diversos indivíduos a partir de sites de compartilhamento de vídeo. Tais imagens foram processadas e selecionadas com base em diferentes aspectos, como idade, sexo, etnia e duração do registro. O resultado do trabalho viabiliza o estudo aprofundado na área de análise de envelhecimento e reconhecimento facial, algo que não era possível com as bases existentes.

Abstract—The main goal of this work is to develop database specific to facial aging, mainly focused on the development of cross-age face recognition methods. To this end, images of many individuals were gathered from video sharing websites. Such images were processed and selected based on different aspects like age, gender, ethnic group and register duration. This database allows in-depth studies in the field of aging analysis and face recognition possible, which was not possible with existing databases.

Keywords-computer vision, aging dataset, facial analysis

I. INTRODUÇÃO

Apesar da grande evolução na área de reconhecimento facial nos últimos anos, o estudo do reconhecimento de faces sob variações causadas pelo envelhecimento ainda caminha a passos curtos. Vários estudos anteriores foram dedicados a resolver um ou mais dos quatro problemas relativos ao reconhecimento facial (i.e. pose, iluminação, expressão e envelhecimento) [1]. O surgimento de novos métodos para detecção facial e extração de características faciais [2], [3] marcam um intenso avanço quanto aos três primeiros problemas citados. Entretanto, a aparência facial de um indivíduo muda de forma razoável com o passar do tempo e os métodos de reconhecimento facial convencionais não são efetivos quando aplicados sob tais circunstâncias. Portanto, a análise das variações faciais causadas pelo processo de envelhecimento é de extrema importância para o aprimoramento destes métodos.

Um dos principais obstáculos no avanço do estudo do reconhecimento facial sob efeitos do envelhecimento é a escassez de bases de imagens de grande porte que atendam as características necessárias [4], explicadas na seção II. Temos alguns exemplos de bases consolidadas [5], [6], [7], mas atender todas as características ainda é um trabalho desafiador. A Internet hoje dispõe de uma vasta quantidade de dados

que podem ser aproveitados, desde conteúdo para leitura a ferramentas de simulação e estimativa de envelhecimento. Dentre tais tipos de dados disponibilizados publicamente, está um de extrema importância na construção de uma base voltada a envelhecimento: imagens faciais de indivíduos em diferentes idades.

Este trabalho teve como objetivo coletar informações e catalogar uma quantidade considerável de imagens de pessoas que registraram o seu processo de envelhecimento através de vídeos e disponibilizaram na Internet. Além disso, buscamos levantar e analisar dados que viabilizem a disponibilização dessas imagens numa nova base voltada especificamente à análise das variações faciais causadas pelo processo de envelhecimento.

II. TRABALHOS RELACIONADOS

Muito já se fez em relação a base de imagens voltadas a envelhecimento. Entretanto, pela dificuldade em coletar uma grande quantidade de imagens de indivíduos voluntários, a maioria dessas bases não atendem todas as características necessárias [4] para o avanço dos estudos na área. Algumas importantes são:

- 1) Distribuição de idade: é importante que a base ofereça cobertura sobre as mais distintas classes de idade, uma vez que se sabe que para algumas dessas as variações faciais extraclasse são diferentes das variações intraclasse. Crianças e adultos, por exemplo, apresentam diferentes tipos de variações entre si. Além disso, em estudos sobre variações faciais mais minuciosos, é importante que as imagens estejam distribuídas de forma densa (i.e., que o intervalo entre uma imagem e outra de um mesmo indivíduo seja pequeno).
- 2) Quantidade e variabilidade de informações: os sistemas relacionados a envelhecimento são geralmente baseados em aprendizado de máquina. Portanto, espera-se que com um grande conjunto de imagens para etapa de treinamento, as técnicas de classificação se tornem mais robustas, ou seja, que o método utilizado mostre bons resultados nos mais distintos cenários.
- 3) Variações do mundo real: pose, expressão, iluminação e vestimentas são fatores importantes, já que é natural que imagens do mundo real contenham muitas variações nesses aspectos.

- 4) Qualidade da imagem: a qualidade de imagem é um ponto importante, pois não é natural que vídeos de indivíduos em instantes consideravelmente distintos apresentem qualidade semelhante. Os dispositivos usados para capturar tais imagens evoluem. Além disso, não é garantido que as imagens de um indivíduo que estejam disponíveis apresentem um padrão de qualidade esperado. Portanto, é importante que os métodos desenvolvidos sejam robustos a tais variações.
- 5) Direitos de distribuição: a depender da técnica usada para coletar as imagens da base e dos propósitos da mesma, distribuir legalmente todas as informações coletadas se torna trabalhoso e, às vezes, impraticável.



Figura 1. Pares de imagens de quatro indivíduos na base de imagens CACD.

A. FG-NET, MORPH

As bases de imagem FG-NET [5] e MORPH [6] são bases antigas, disponibilizadas publicamente e muito utilizadas na literatura. Entretanto, tais bases contém poucas imagens de um mesmo indivíduo. Além disso, na FG-NET a quantidade de imagens é pequena e a MORPH não contém praticamente nenhuma variação do mundo real. Tais características implicaram em um estudo mais aprofundado sobre os principais desafios para construção de uma base voltada a envelhecimento.

B. VADANA

A base de imagens [4] tenta suprir a maioria das características que as duas bases citadas anteriormente não contemplaram, introduzindo um banco com maior cobertura de idade, variações de pose, iluminação e expressão facial. O seu grande diferencial está na grande quantidade de imagens por indivíduo. Entretanto, o intervalo entre imagens de um mesmo indivíduo ainda é grande e a quantidade total de imagens é pequena.

C. CACD

A *Cross-Age Celebrity Dataset* (CACD) foi construída a partir de mineração de imagens de celebridades do Google¹ e da necessidade de um banco robusto para validar um método de reconhecimento facial invariante a envelhecimento [7]. Devido a grande variabilidade de imagens de celebridades no Google, a CACD contém uma grande quantidade de indivíduos e de imagens, e os registros têm duração média razoável. Entretanto, a quantidade média de imagens por indivíduo ainda é pequena para o estudo aprofundado das variações causadas pelo envelhecimento. A Figura 1 mostra alguns pares de imagens de celebridades do banco.

A nossa base tem como objetivo, além de preencher razoavelmente as características citadas, o diferencial de fornecer imagens de um mesmo indivíduo tal que o intervalo entre duas imagens consecutivas seja sempre pequeno, visando os estudos das variações faciais causadas pelo envelhecimento de forma contínua.

¹<http://google.com>

III. PROCESSO

Na primeira etapa do trabalho tivemos a tarefa de reunir a maior quantidade de imagens pertinentes ao problema no site de compartilhamento de vídeos Youtube². O mesmo foi adotado como fonte de coleta porque concentra uma grande quantidade de vídeos de usuários que registraram e disponibilizaram o seu processo de envelhecimento.

O período de trabalho pode ser dividido em três etapas, de acordo com a Figura 2:



Figura 2. Fluxograma do processo de construção da base.

A. Coleta de vídeos

Na primeira e mais longa etapa, vídeos do Youtube que registravam o processo de envelhecimento de pessoas foram coletados. Tais vídeos foram escolhidos com base na variação de idade e de fisionomia, além da qualidade das imagens e duração dos registros.

É importante ressaltar que os vídeos coletados devem registrar o processo de envelhecimento facial do indivíduo. Logo, um critério importante para seleção destes foi a presença predominante da face na imagem.

Além dos arquivos de vídeo, as referências aos vídeos originais no Youtube também foram armazenadas.

B. Levantamento de informações sobre os vídeos

É importante que a base contenha informações sobre os indivíduos presentes. Além das imagens estáticas, os seguintes dados foram levantados:

²<http://youtube.com>

- 1) Idade estimada do indivíduo no começo do registro;
- 2) Duração do registro em dias;
- 3) Etnia;
- 4) Sexo.

C. Extração e processamento das imagens

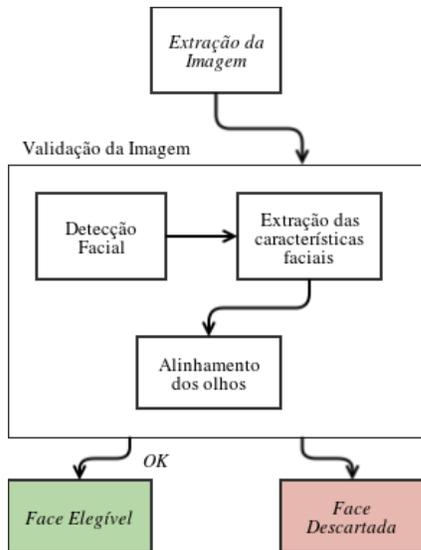


Figura 3. Fluxograma da etapa de extração e processamento de imagens.

Na terceira etapa, trabalhamos na extração e no processamento de imagens estáticas extraídas dos vídeos coletados. A Figura 3 mostra uma visão geral do processo. Tais imagens foram extraídas de forma automática com a ferramenta FFmpeg³, através do seu filtro de detecção de mudança de cena, configurado com um limiar de 0,05. Tal limiar foi escolhido com base em testes prévios num pequeno conjunto de vídeos. Essa escolha diminuiu também os riscos de que imagens duplicadas fossem extraídas. Entretanto, nesta etapa do processo, podem ainda haver imagens sem face ou uma quantidade pequena de duplicatas dentre as imagens extraídas. A Figura 4(a) mostra um exemplo de imagem extraída.

Em seguida, as imagens foram submetidas ao processo de validação de imagem. Uma imagem é válida se contiver uma *face elegível*. O framework de detecção de objetos proposto por Viola-Jones [8], presente na biblioteca OpenCV [9], foi escolhido para o processo de detecção facial. A Figura 4(b) mostra um exemplo de resultado do processo de detecção facial. Além da face, extraímos também oito localizações de características faciais utilizando a biblioteca Flandmark [10]. A Figura 4(c) mostra um exemplo de todas as localizações de características faciais desenhadas.

O critério usado para validação da imagem foi a detecção de uma única face quando aplicado o algoritmo de detecção facial. O classificador em cascata utilizado foi o de faces frontais previamente treinado e distribuído com a biblioteca

OpenCV. Como os vídeos coletados são vídeos casuais publicados na Internet, e não registros feitos especialmente para análises faciais, faces com as mais distintas poses não-frontais estão entre as imagens extraídas. O fato do método de Viola e Jones e o classificador em questão não funcionarem bem com faces rotacionadas ajudou no processo de descarte dessas imagens.

As imagens também são rotacionadas de forma que os centros dos olhos fiquem alinhados horizontalmente e então são cortadas de forma que a imagem final contenha somente a face frontal. A imagem final é um quadrado com tamanho igual a quatro vezes a distância entre os centros dos olhos centralizada no nariz do indivíduo [7]. Quando tal quadrado excede os limites da imagem original, a própria região obtida com o algoritmo de detecção facial é utilizada. Uma aproximação boa do centro de um olho é definida como o ponto médio entre os dois extremos do mesmo, já obtidos com o Flandmark. A Figura 4(d) mostra um exemplo do resultado final do processo.

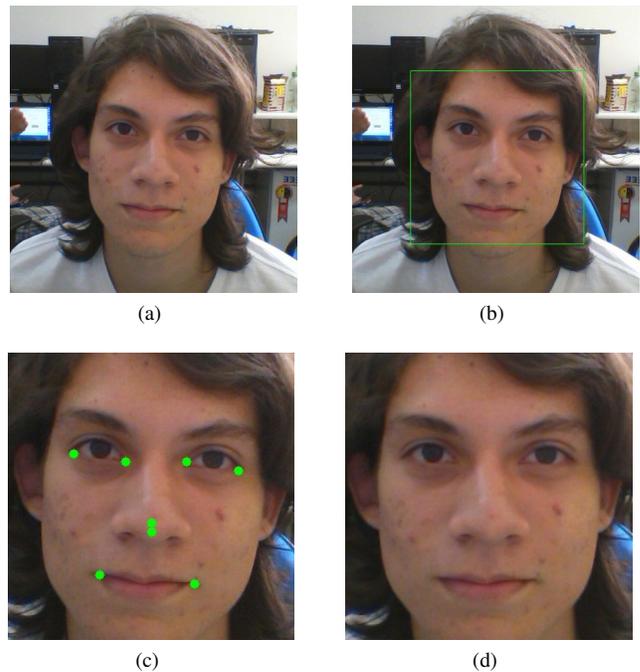


Figura 4. Exemplo dos resultados da extração de imagem, (b) detecção facial, (c) extração de características faciais e (d) alinhamento dos olhos.

Além disso, para cada face i elegível, uma estimativa de idade a_i para o indivíduo naquela imagem é calculada conforme a equação (1). Tal cálculo é feito a partir da idade estimada do indivíduo no começo do registro (a_0), da duração d do registro (em dias) e da imagem extraída q_i em que a face aparece. É considerado como começo do registro o primeiro quadro com uma face extraída e m é a quantidade de faces extraídas.

$$\Delta t_i = q_i / m \cdot d$$

$$a_i = \frac{\Delta t_i}{365} + a_0 \quad (1)$$

³<http://ffmpeg.com/>

onde Δt_i é o intervalo em dias entre a primeira imagem e a imagem q_i , e a_i é a idade estimada do indivíduo na imagem q_i .

IV. RESULTADOS

No total, foram coletados 113 vídeos de usuários distintos. A duração real dos registros varia de alguns meses até dezenas de anos, como pode ser observado na Figura 5. Após a coleta, 62 destes vídeos foram descartados manualmente com base nos critérios de predominância facial, duração do registro, variação de pose, sexo e disponibilidade. Foram armazenadas, por fim, as seguintes informações desses vídeos:

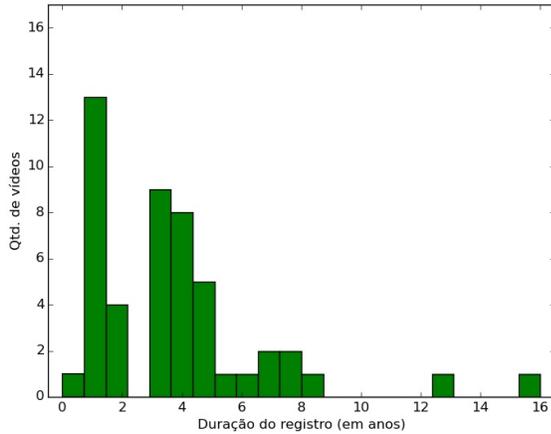


Figura 5. Distribuição dos registros quanto à duração.

- 1) URL e nome do registro no YouTube;
- 2) Duração real do registro (tempo passado entre a captura da primeira e da última imagem);
- 3) Idade estimada do indivíduo no começo do processo de registro;
- 4) URL para obtenção do arquivo de vídeo;
- 5) Sexo do indivíduo;
- 6) Etnia do indivíduo.

Dos vídeos restantes foram extraídas 56038 imagens estáticas. Destas, 10,4% foram descartadas na etapa de processamento das imagens (i.e imagens sem *faces elegíveis*). A Tabela I descreve a proporção dos indivíduos quanto suas características, após cada etapa. Devido a falta de vídeos no YouTube que atendam as características citadas na seção III-A, é possível notar a escassez de indivíduos negros e asiáticos, assim como do sexo feminino. A Figura 6 mostra a distribuição das faces selecionadas quanto à idade e ao sexo dos indivíduos.

Além das imagens e dos seus dados, também são geradas como resultado as posições das oito características faciais para todas as faces elegíveis. A Figura 7 mostra algumas imagens de faces elegíveis de um mesmo indivíduo em ordem cronológica. O vídeo⁴ do indivíduo mostrado está sob a

⁴<https://www.youtube.com/watch?v=XjSfaRzbAHc>

licença *Creative Commons* e, portanto, suas imagens podem ser utilizadas para qualquer fim.

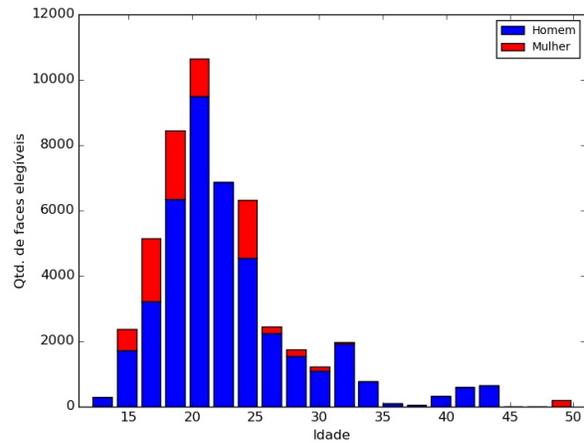


Figura 6. Distribuição das faces elegíveis quanto à idade e ao sexo.

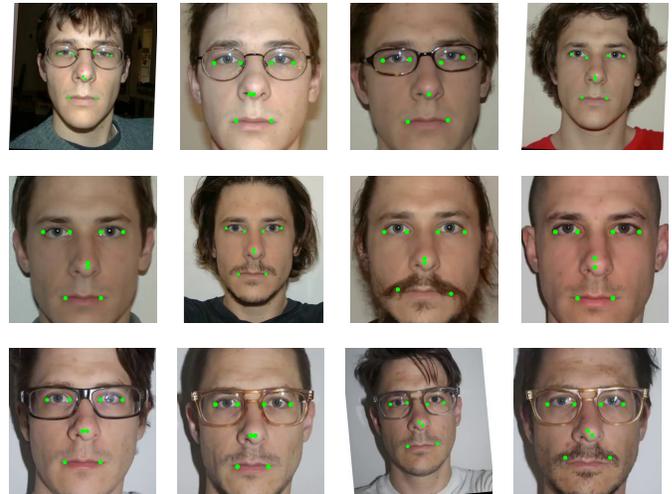


Figura 7. Doze faces elegíveis de um mesmo indivíduo, com as localizações de suas características faciais desenhadas.

A Tabela II mostra alguns dados comparativos da nossa base e de outras bases citadas. Percebe-se que o grande diferencial está na densidade da base (i.e., há uma grande quantidade de imagens de um mesmo indivíduo e intervalo entre essas imagens é ínfimo). Isso possibilita uma análise minuciosa sobre o processo de envelhecimento.

V. EXPERIMENTOS

Dois experimentos foram conduzidos com os dados presentes na base de imagens desenvolvida. Ambos utilizaram o How-Old.net⁵ para estimar a idade de alguns indivíduos a partir de alguns subconjuntos de faces elegíveis dos mesmos.

⁵<http://how-old.net>

Tabela I
DADOS QUANTITATIVOS E QUALITATIVOS DA BASE DE IMAGENS.

	Homens	Mulheres	Branco	Negros	Asiáticos	Total
Vídeos	41	8	45	2	2	49
Imagens extraídas	46766	9272	50539	828	4671	56038
Faces elegíveis	41856	8374	45032	793	4405	50230
Duração média dos registros (em anos)	3,8	3,5	3,9	1,0	4,0	3,8
Duração total dos registros (em anos)	156,2	28,0	174,2	2,0	8,0	184,2

Tabela II
COMPARAÇÃO ENTRE AS BASES DE IMAGENS.

Base de Imagens	# de imgs.	# de indiv.	# de imgs./indiv.	Duração média
FGNET [5]	1002	82	12,2	0-45
MORPH [6]	55134	13618	4,1	0-5
CACD [7]	163446	2000	81,7	0-10
VADANA [4]	2298	43	53	0-28
Nossa base	50230	49	1025,1	0-4

O How-Old.net é um serviço web da Microsoft⁶ que estima a idade de indivíduos a partir de suas imagens faciais. O website oferece uma interface simples para que o usuário possa enviar sua própria imagem. As estimativas são feitas com base em uma API de aprendizado de máquina especializada em faces do projeto Oxford⁷, conduzido pelo grupo de pesquisa da própria empresa.

A. Experimento Individual

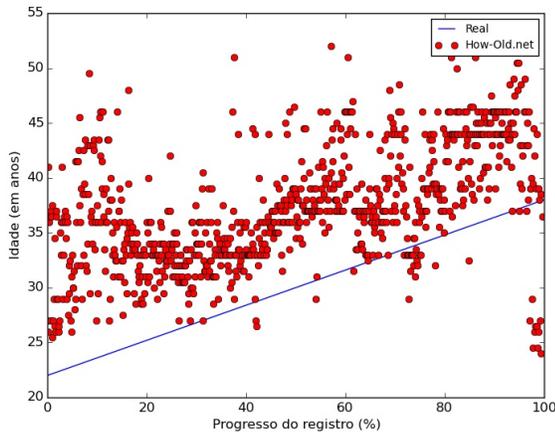


Figura 8. Estimativas de idade para cada face elegível do registro.

No primeiro experimento, o maior dos vídeos, levando em conta a duração do registro em anos, foi escolhido (o mesmo da Figura 7). O usuário registrou seu processo de envelhecimento diariamente por 16 anos. Todas as faces elegíveis extraídas do vídeo foram submetidas ao How-Old.net e as estimativas de idade para cada uma delas foi armazenada. A relação entre a idade real do usuário e a estimativa de idade dada pelo How-Old.net em cada instante do registro pode ser

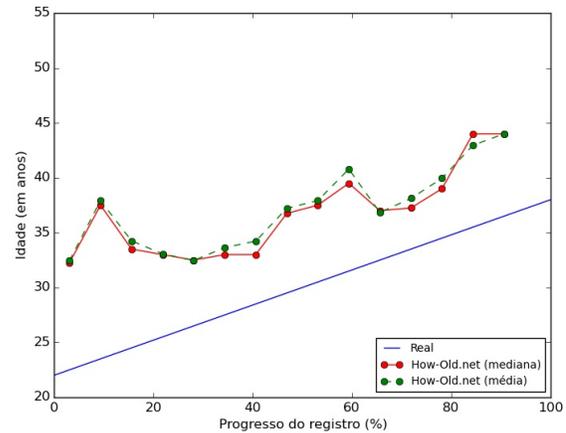


Figura 9. Análise temporal das estimativas do registro.

vista na Figura 8. Na Figura 9 estão representadas as medianas e as médias de cada intervalo de um ano do registro. É possível notar que em pequenos intervalos de tempo existem muitos outliers e o erro é grande. Entretanto, as estimativas começam a apresentar uma tendência de crescimento ligeiramente semelhantes ao crescimento da idade real quando classificadas em intervalos de um ano. Tais observações evidenciam que o método usado pelo How-Old.net não funciona bem com dados contínuos (i.e., uma grande quantidade de imagens com intervalos de tempo curtos). A nossa base de imagens viabiliza, portanto, o estudo das variações faciais causadas pelo envelhecimento em menor escala.

B. Experimento Geral

Como é impraticável enviar todas as imagens da base ao How-Old.net, no segundo experimento, para todo indivíduo da base, amostras de dez imagens foram coletadas para cada idade presente no registro. Para cada amostra, essas imagens foram enviadas ao How-Old.net e os resultados armazenados.

Na Figura 10 estão representadas as medianas das amostras

⁶<http://microsoft.com>

⁷<http://www.projectoxford.ai/>

dos indivíduos para cada idade. Já que as imagens são de vídeos do YouTube e estão sujeitas a variações incomuns, as medianas foram usadas com o intuito de minimizar os efeitos dos erros pontuais sobre o resultado final. Percebe-se que o How-Old.net tende a estimar idades maiores que a idade real. Além disso, é possível notar que os palpites, apesar dos erros absolutos, acompanham de forma razoável a curva de envelhecimento. Já a Figura 11 mostra as estimativas para todas as amostras de todos os indivíduos. É possível perceber que a ocorrência de outliers neste experimento é menor que no experimento individual.

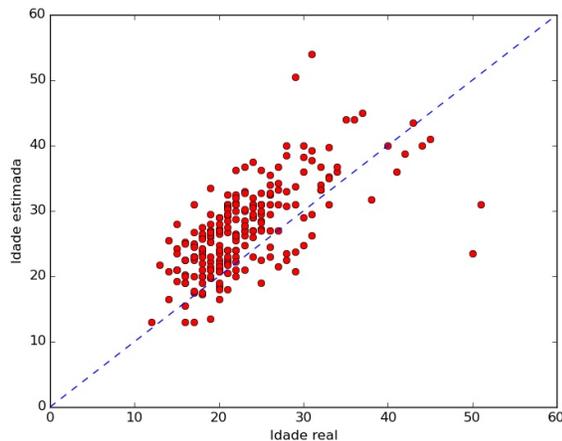


Figura 10. Medianas das amostras de todos os indivíduos da base para cada idade.

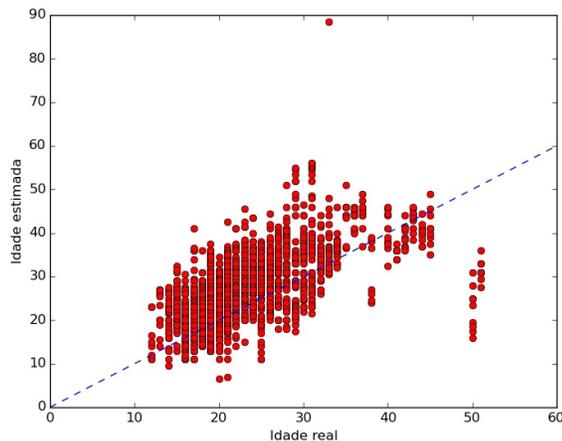


Figura 11. Estimativas das amostras de todos os indivíduos da base para cada idade.

VI. CONCLUSÃO

As análises apresentadas mostram que o estudo das variações faciais causadas pelo envelhecimento a partir das

bases que existem atualmente não é suficiente para sanar todos os problemas associados a estimação e simulação de envelhecimento. Isto se deve, principalmente, ao fato das imagens de um mesmo indivíduo nessas bases estarem separadas por um largo intervalo de tempo, i.e. os métodos estudados são desenvolvidos sem levar em conta dados contínuos. Portanto, uma base com as características como as da base proposta é importante para avaliar a robustez dos métodos.

O fato de estarmos limitados ao YouTube implica na baixíssima variabilidade da nossa base. Praticamente não existem vídeos de mulheres ou de homens negros e asiáticos no sistema que registrem o seu processo de envelhecimento, por exemplo. Contudo, o fato desta base usar como principal fonte a Internet permite que a mesma cresça e varie com o passar do tempo, possibilitando que tais pontos possam ser melhorados no futuro.

Quando o trabalho estiver encerrado, todos os dados e códigos para geração da base serão disponibilizados. Como evidenciado na Figura 5, na Figura 6 e na tabela I, ainda existem várias lacunas a serem preenchidas pela base em desenvolvimento. Como trabalhos futuros pretendemos coletar vídeos com maior variabilidade de características e desenvolver mais experimentos que validem a importância da base e dos dados presentes na Internet para o estudo do envelhecimento.

AGRADECIMENTOS

Os autores gostariam de agradecer à UFBA e ao CNPq pelo apoio financeiro.

REFERÊNCIAS

- [1] U. P. A. K. Jain, B. Klare, "Face matching and retrieval in forensics applications," in *IEEE Multimedia*, vol. 19. IEEE, 2012, p. 20.
- [2] F. D. Xuehan Xiong, "Supervised descent method and its applications to face alignment," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2013.
- [3] Z. L. F. Zhou, J. Br, "Exemplar-based graph matching for robust facial landmark localization," 2013.
- [4] G. Somanath, M. Rohith, and C. Kambhamettu, "Vadana: A dense dataset for facial image analysis," in *Computer Vision Workshops (ICCV Workshops), 2011 IEEE International Conference on*, Nov 2011, pp. 2175–2182.
- [5] Face and Gesture Recognition Working group, "Fg-net aging database," 2000.
- [6] K. Ricanek and T. Tesafaye, "Morph: a longitudinal image database of normal adult age-progression," in *Automatic Face and Gesture Recognition, 2006. FGR 2006. 7th International Conference on*, April 2006, pp. 341–345.
- [7] W. H. H. Bor-Chun Chen, Chu-Song Chen, "Face recognition using cross-age reference coding with cross-age celebrity dataset," in *IEEE Transactions on Multimedia*. IEEE, 2015.
- [8] M. J. P. Viola, "Robust real-time face detection," *International Journal of Computer Vision*, vol. 57, pp. 137–154, 2004.
- [9] G. Bradski, "The OpenCV Library," *Dr. Dobb's Journal of Software Tools*, 2000.
- [10] M. Uříčář, V. Franc, and V. Hlaváč, "Detector of facial landmarks learned by the structured output SVM," in *VISAPP '12: Proceedings of the 7th International Conference on Computer Vision Theory and Applications*, G. Csurka and J. Braz, Eds., vol. 1. Portugal: SciTePress — Science and Technology Publications, February 2012, pp. 547–556.