

UNICAMP - Universidade Estadual de  
Campinas  
MO444 - Aprendizado de Máquina  
Trabalho Individual 2

David Burth Kurka\*

2 de outubro de 2013

## 1 Introdução

Este trabalho busca estudar diferentes alternativas para o problema de reconhecimento de faces. Dispondo de um conjunto base com 3214 imagens de rosto de 1203 indivíduos distintos, procura-se identificar a identidade de novas fotos apresentadas, a partir de comparação de similaridade. Para o teste de correção do algoritmo, são usadas 200 novas imagens dos indivíduos que compõem o conjunto base e a taxa de acerto é avaliada.

As técnicas utilizadas diferem na representação das imagens processadas e nas métricas de similaridade comparando as fotos novas com as fotos do banco de dados. A especificação e comparação da eficiência das várias técnicas será descrita nas sessões seguintes.

## 2 Pré processamento

Para que seja possível haver comparação de similaridade entre fotos, um pré processamento é necessário, a fim de encontrar padrões na representação das imagens trabalhadas. Era conhecido a priori, de todas as imagens trabalhadas, a posição (pixels) dos olhos na imagem. Desse modo, todas as imagens analisadas passaram pela seguinte sequência de operações:

- **Redimensionamento:** encontrado um valor médio para a distância entre os olhos das imagens (70 pixels), todas as imagens foram redimensionadas para apresentar esta distância padrão;
- **Rotação:** as imagens que não apresentavam as coordenadas de olho alinhadas horizontalmente, foram rotacionadas, de forma a corrigir o desvio;
- **Recorte:** após redimensionada e rotacionada, foi feito um recorte retangular, de forma a centralizar o rosto e eliminar o fundo.

---

\*RA: 070589, email: david.kurka@gmail.com

Após esse processamento, todas as imagens trabalhadas passaram a ter dimensão 151X196 pixels, com o rosto alinhado em seu centros.

### 3 Experimento 1 - Distância entre imagens

A técnica mais imediata para o reconhecimento de novas imagens é o cálculo pixel a pixel de similaridade entre duas imagens. A expectativa é que, ao se apresentar uma nova foto, a imagem do banco cujos pixels estejam mais próximos seja do mesmo indivíduo que procura-se identificar.

Para calcular a similaridade entre duas imagens, calcula-se a distância euclidiana entre as duas imagens, sendo a imagem com menor distância a mais semelhante. Como as imagens passaram por pré-processamento e portanto estão padronizadas, espera-se que pixels com a mesma localização em imagens diferentes correspondam às mesmas posições do rosto representado, sendo então possível a identificação e diferenciação de indivíduos com essa técnica.

Entretanto, este método possui um alto custo, já que cada imagem possui 29596 pixels que devem ser comparados ponto a ponto, para cada uma das 3214 imagens do banco. Esse fator é um dos principais limitadores do método; para a classificação cujos resultados serão descritos a seguir, o processo levou diversas horas, enquanto os outros processamentos eram realizados em segundos ou no máximo minutos.

#### 3.1 Resultados

Das 200 imagens apresentadas, **159** foram identificadas corretamente. Isto corresponde a uma taxa de **79,5%** de acerto.

### 4 Experimento 2 - Eigenfaces (PCA)

Uma alternativa interessante para o cálculo das distâncias entre as imagens é a redução da dimensão das mesmas, antes da comparação. Uma forma de fazer isso é a partir da análise de componentes principais (PCA).

Os componentes principais são encontrados a partir do cálculo dos autovetores da matriz de covariância dos pixels das imagens do banco de dados. Porém, como as imagens possuem dimensão na ordem de dezenas de milhares de elementos, sua matriz de covariância possui centenas de milhares de elementos, tornando o seu cálculo inviável.

Para contornar essa limitação utilizou-se a técnica apresentada por Turk e Pentland [2], que explora o fato do número de imagens ( $M$ ) ser menor que o número de pixels ( $n$ ), limitando a quantidade de autovalores não-nulos da matriz de covariância a no máximo  $M$  (posto da matriz de dados). Dessa forma, pode-se encontrar primeiro os autovetores associados à matriz de covariância das imagens (matriz  $MXM$ , tratável) e então multiplica-la pela matriz  $NXM$  da base de dados, encontrando os  $M$  autovetores da matriz de covariância dos pixels.

O passo seguinte é definir o número de componentes principais que serão utilizados. Devem ser escolhidos os autovetores associados aos autovalores com maiores valores, sabendo-se que o erro da projeção é função da soma dos autovalores não escolhidos. Para analisar o impacto da escolha, foi analisado o

gráfico da soma acumulada dos autovalores (figura 3). Nele vemos que, percentualmente, a maior concentração dos valores estão nos primeiros elementos, havendo diversos autovalores com valores não significativos. A partir da análise foram escolhidos os 1000 primeiros autovetores para formar a base de componentes principais, cujos autovalores associados correspondem à mais de 99% da soma de todos os elementos.

É interessante notar que, apesar da diminuição de dimensionalidade produzir erros nas representações das imagens, ela também é capaz de criar generalizações e aproximações, evitando um processo de *overfitting*. Dessa forma, no processo de redução de dimensão, regiões pouco relevantes das imagens podem se perder na representação, de forma que a comparação de similaridade se concentre em regiões mais decisivas para a classificação.

Outra característica interessante do método é que os autovetores calculados possuem as mesmas dimensões de uma imagem e podem ser exibidos graficamente, como mostra a figura 2. Essas imagens foram chamadas de “*eigenfaces*” por [2] e correspondem à imagens “bases” das quais qualquer outra face pode ser derivada, como combinação.

O processo de cálculo de autovetores resulta na obtenção de uma matriz de transformação  $W_{pca}$  de dimensões  $n$  por  $R$  (onde  $R$  é a dimensão de redução, no caso 1000). Ao multiplicar a transposta de  $W_{pca}$  pela base, obtemos a base reduzida, de dimensão  $R$  por  $M$ , representando as  $M$  imagens do conjunto. A partir da base reduzida podemos fazer o cálculo de similaridade com novas imagens também reduzidas, como no experimento anterior, porém com velocidade muito maior, devido à diminuição abrupta de dimensionalidade.

## 4.1 Resultados

Das 200 imagens apresentadas, **126** foram identificadas corretamente. Isto corresponde a uma taxa de **63%** de acerto.

## 5 Experimento 3 - Fisherfaces (LDA)

Uma das principais desvantagens da redução utilizando componentes principais é que o método procura bases que possam representar a maior variação possível de imagens, sem levar em conta características específicas dos indivíduos da base. Dessa forma, o PCA se torna limitado quando, em imagens diferentes de um mesmo indivíduo, pequenas mudanças como, diferença no posicionamento do rosto ou na luminosidade da foto, separam muito as projeções, não sendo possível encontrar similaridade entre as fotos.

Como alternativa ao PCA, portanto, é apresentada a técnica de *linear discriminant analysis* que é capaz de levar em conta informações a respeito das classes das imagens, sendo a classe definida pelo conjunto de fotos de um mesmo indivíduo. Na base utilizada, existem 1203 classes, correspondendo à 1203 indivíduos distintos.

O método consiste em encontrar uma projeção que ao mesmo tempo minimiza o espalhamento interno ( $S_w$ ) e maximiza o espalhamento entre as classes ( $S_b$ ). O processo é análogo ao método do PCA, onde a partir da matriz de autovetores (agora do problema generalizado  $S_b w = \lambda S_w w$ ) encontra-se a matriz de projeção  $W_{lda} \in \mathbb{R}^{R \times n}$ , capaz de reduzir a dimensionalidade das imagens.

Porém, há uma particularidade para o cálculo da inversa de  $S_w$ , como descrito em [1], pois a mesma é singular em  $\mathbb{R}^n$ . A solução consiste em primeiro reduzir a dimensionalidade da base para  $m - c$  (posto de  $S_w$ , onde  $c$  é o número de classes do conjunto) utilizando PCA e então calcular o LDA da base. A dimensão resultante escolhida da projeção é igual à 1202 ( $c - 1$ ) por  $m$ , que é o mínimo valor de redução do LDA e o valor utilizado nas referências estudadas.

Uma característica interessante desse método é que, assim como no PCA, os autovetores podem ser visualizados graficamente e agora são chamados de “*fisherfaces*” (em homenagem à Ronald Fisher, que introduziu a técnica) e correspondem à “bases” aperfeiçoadas para distinguir classes. A projeção de algumas das obtidas podem ser vistas na figura 4.

## 5.1 Resultados

Das 200 imagens apresentadas, apenas **49** foram identificadas corretamente, o que corresponde a uma taxa de **24.5%** de acerto.

## 6 Experimento 4 - PCA + LDA

Como a base possui muitas classes com poucos elementos (a base possui 1203 indivíduos, onde cada individuo possui em média 2.67 fotos, havendo vários indivíduos com apenas uma foto), o LDA pode apresentar falhas quando aplicado sozinho. Diante disso, pode ser vantajoso primeiro efetuar a redução de dimensionalidade na base, utilizando PCA e então aplicar o LDA, de forma a tratar as particularidades de cada classe. Esta técnica é apresentada, entre outros, por Zhao et Al. [3].

Dessa forma, primeiro calculamos a matriz de transformação  $W_{pca}$  ( $n \times R$ ) e então calculamos o PCA na base reduzida, gerando a matriz  $W_{lda}$  ( $R' \times R$ ). A combinação das matrizes resulta na matriz  $W$ , que permite a transformação final da base e das imagens a ser analisadas.

### 6.1 Resultados

Das 200 imagens apresentadas, **142** foram identificadas corretamente. Isto corresponde a uma taxa de **71%** de acerto.

## 7 Experimento 5 - Outras métricas de distância

Todos os experimentos acima usaram distância euclidiana como métrica de similaridade, para fazer comparações e encontrar imagens próximas. Nesta etapa, utilizaremos diversas outras medidas e comparar com o resultado obtido para o reconhecimento de faces utilizando **PCA+LDA** (experimento 5).

As métricas utilizadas e comparadas foram: distância de Manhattan, norma infinita (distâncias de minkowski), cosseno, Canberra, Bray-Curtis e correlação.<sup>1</sup>

Os resultados obtidos foram (em ordem de desempenho):

<sup>1</sup>Formulas e descrição das medidas utilizadas podem ser encontradas em <http://reference.wolfram.com/mathematica/guide/DistanceAndSimilarityMeasures.html>

| Método            | Acertos (em 200) | Porcentagem |
|-------------------|------------------|-------------|
| Manhattan         | 118              | 59%         |
| Norma Máxima      | 141              | 70,5%       |
| <b>Euclidiana</b> | <b>142</b>       | <b>71%</b>  |
| Canberra          | 147              | 73,5%       |
| Norma Máxima      | 141              | 70,5%       |
| Bray-Curtis       | 183              | 91,5%       |
| Cosseno           | 189              | 94,5%       |
| Correlação        | 189              | 94,5%       |

## 8 Comparação de resultados e Conclusão

Ao analisar os resultados obtidos nos experimentos descritos acima, nota-se algumas características que merecem ser destacadas.

É notável que a simples comparação de similaridade das imagens, sem redução de dimensionalidade (experimento 1) obteve o melhor desempenho de acerto no reconhecimento de faces, utilizando distância euclidiana alto desempenho. Entretanto, o tempo de processamento dessa operação é um fator a ser considerado e provavelmente um limitante.

Com a diminuição de dimensionalidade, a partir de análise de componentes principais (pca-eigenfaces), a eficiência do método diminui, porém ainda atinge níveis aceitáveis (acima de 60% de acerto). Nessa situação, apesar da queda na confiabilidade da classificação, a velocidade da computação da distância torna essa técnica atraente, devido ao seu baixo custo.

Entretanto, quando se utiliza puramente o LDA para reduzir a dimensionalidade, o resultado é foi muito negativo. Uma possível explicação para esse efeito pode ser o fato do conjunto base possuir muitos indivíduos distintos (classes), com poucas imagens cada.

Mas, ao se aplicar o método LDA em base já reduzida com PCA (como feito no experimento 4), observa-se melhor desempenho comparado com as técnicas separadas. Combinado com o baixo custo de processamento desta técnica, esta técnica parece ser a melhor alternativa vista.

A exploração de outras medidas de distância (ou similaridade), mostra que existem métricas mais interessantes que a euclidiana. Destaque para a distância cosseno, cujo cálculo é pouco custoso e foi capaz de obter quase 95% de acerto!

Com esse trabalho, concluímos que o problema de classificação de faces é um problema viável, possuindo técnicas, como as vistas acima, capazes de reconhecer faces com grandes taxas de acerto e em tempo reduzido.

## Referências

- [1] P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman. Eigenfaces vs. fisherfaces: Recognition using class specific linear projection. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 19(7):711–720, 1997.
- [2] M. Turk and A. Pentland. Eigenfaces for recognition. *Journal of cognitive neuroscience*, 3(1):71–86, 1991.

- [3] W. Zhao, A. Krishnaswamy, R. Chellappa, D. L. Swets, and J. Weng. Discriminant analysis of principal components for face recognition. In *Face Recognition*, pages 73–85. Springer, 1998.



Figura 1: Imagem com média de todos os pixels do conjunto base de imagens. Como o conjunto contém imagens tanto masculinas e femininas, nota-se que a parte mais definida da média é a região dos olhos, por ser uma parte em comum em todos os indivíduos.

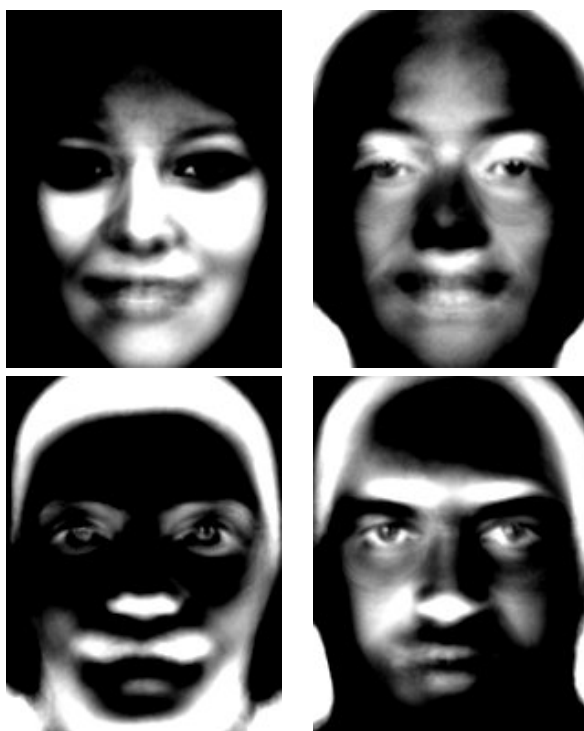


Figura 2: “*Eigenfaces*” da base de imagens. Essas imagens são usadas como base para, quando combinadas, formar imagens reais de rostos. Note que cada imagem pode destacar características, como detalhes de um rosto feminino, cabelo, barba, óculos.

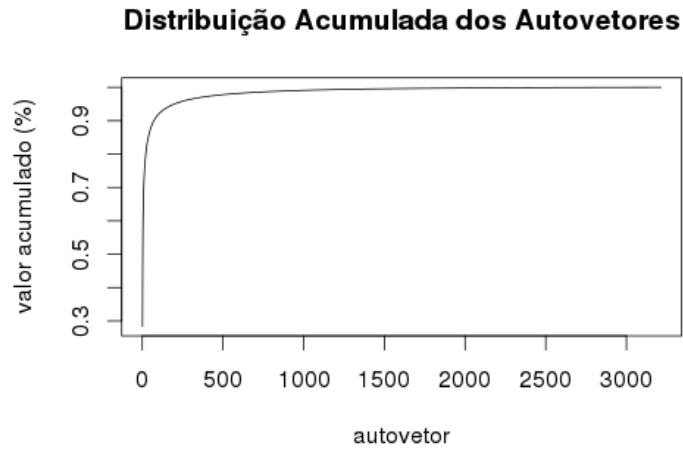


Figura 3: Soma acumulada dos autovalores associados à matriz de covariância das imagens da base, ordenados decrescentemente. Note que os autovalores iniciais possuem valores significativos, enquanto grande parte possui valores baixos, tendendo a zero (região estagnada do gráfico). Os 100 maiores autovalores correspondem a mais de 90% da soma total de todos os autovalores; os 1000 maiores autovalores correspondem a mais de 99%

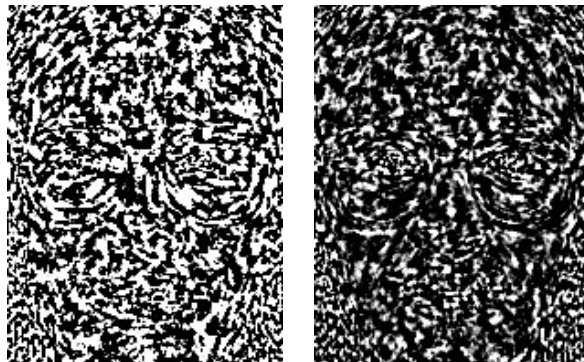


Figura 4: “*Fisherfaces*”. Apesar da semelhança com obras do impressionista Van Gogh, as *fisherface* possuem informações específicas de indivíduos, possibilitando a distinção de classes. Entretanto, pelo grande número de classes do conjunto, não é fácil perceber padrões e formas bem delimitadas nas imagens obtidas. Em todo o caso, a forma das imagens se assemelha com resultados mostrados em [3] para o mesmo caso.