Pattern Recognition: Coursework 1

Chan, Bingjie
bc1714
CID: 00979071

Ng, Zhi Long Benjamin
bn514
CID: 00986013

Abstract

The ABSTRACT is to be in fully-justified italicized text, at the top of the left-hand column, below the author and affiliation information. Use the word "Abstract" as the title, in 12-point Times, boldface type, centered relative to the column, initially capitalized. The abstract is to be in 10-point, single-spaced type. Leave two blank lines after the Abstract, then begin the main text. Look at previous CVPR abstracts to get a feel for style and length.

1. Introduction

From security purposes such as criminal identification by law enforcement [3] to entertainment such as automatic face image tagging on social media, machine based face recognition has a plethora of applications. Since the first Automatic Face Recognition system developed by Kanade in 1973, technology in face recognition has shifted from simply detecting and determining relationships among individual facial features of subjects to linear projection appearance-based methods such as Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA). [5]

To develop a real-time system, it is essential that the recognition algorithm is accurate, fast and simple. It is by these metrics of accuracy, efficiency and simplicity that this paper seeks to implement and evaluate both PCA and LDA methods in face recognition.

Chapter 3 elaborates on the pre-processing of the given image data in face.mat, and while Chapter 3.1 details maximising the efficiency and simplicity of PCA as a feature extraction method, Chapter IV expounds on improving the accuracy of PCA for face recognition. (Add in Chapter V and VI later)

2. Data Pre-Processing

To start off, the given face.mat contains a matrix $\boldsymbol{X}$ (2576x520) and a vector $\boldsymbol{l}$ (1x520). The columns of $\boldsymbol{X}$ represent each of the 520 face samples, with each sample being a 56 pixels by 46 pixels image. The elements of $\boldsymbol{l}$ then show 52 individuals (or classes), each with 10 sample images taken with varying backgrounds, clothing and angles. An example of a class is shown in Figure 1.

Figure 1: Original Images of Class 1

3. Eigenfaces

Face recognition systems can be simplified as a template matching problem where given unseen data, the system attempts a 'best match' with a pre-existing data collection. To execute such a comparison in vector spaces, an image of size W by H pixels can be treated as a vector in a subspace of dimensionality $D = W \times H$. Since face images are similar, their data points are not entirely randomly distributed in this space and hence can be represented by a lower dimensionality subspace.

PCA characterises the variation in a collection of faces as 'features' and thus encodes relevant information more efficiently. These features are represented by the eigenvectors (or eigenfaces) of the covariance matrix of training data. Consequently, original images can be reconstructed as a linear combination of these eigenfaces. By choosing the best M eigenvectors that account for the distribution of face images within the original image space D, PCA is a form of dimensionality reduction since a smaller set of basis vectors are used to represent the original data. [4]

3.1. Computation of Eigenfaces

Given face images $\mathbf{I}_1, \mathbf{I}_2, ...\mathbf{I}_N$ from the training set partitioned in Section 2 where N is the number of images in the set, it is pivotal to first normalise $\mathbf{I}_n$ for scale, orientation and translation and then represent it as a vector $\mathbf{x}_n$. The mean face $\bar{\mathbf{x}} = \frac{1}{N}\sum_{n=1}^{N}\mathbf{x}_n$ is calculated and shown in Figure 2.

Figure 2: Mean Face

Each training face differs from the mean face by vector $\phi_n = \mathbf{x}_n - \bar{\mathbf{x}}$ and the normalised training faces form the columns of matrix $\boldsymbol{A}$.

$$\boldsymbol{A} = [\boldsymbol{\phi}_1, \boldsymbol{\phi}_2, ..., \boldsymbol{\phi}_N] \in \boldsymbol{R}^{D \times N} \tag{1}$$

The eigenvalues λ_m and corresponding eigenvectors $\boldsymbol{u}_m$ are then obtained by performing eigendecomposition on the covariance matrix

$$\boldsymbol{S} = \frac{1}{N}\boldsymbol{A}\boldsymbol{A}^T \in \boldsymbol{R}^{D \times D} \tag{2}$$

Since the eigenfaces with the largest eigenvalues account for the greatest data variances, the largest M eigenfaces and their corresponding eigenvalues are chosen to create an effective lower dimensional subspace of eigenfaces (or 'face space') that can best represent the original images. As each image location varies in contribution to each eigenface, the eigenfaces below are represented as ghostly faces. This can be seen by the first 5 eigenfaces corresponding to the largest 5 eigenvalues from the training set in Figure 3 below.

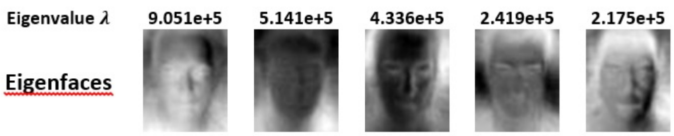

Figure 3: The 'Best' 5 Eigenvalues and Eigenfaces

When the eigenvalues are sorted by magnitude as shown in Figure 6, it is clear that the first 50 eigenvalues contribute the most to the variability of the data and as seen by the 'best' 5 eigenfaces in Figure 3, these have vaguely recognisable human features that people can identify. In contrast, eigenvectors with very low eigenvalues as shown in Figure 4 vaguely forms recognisable human features.

Figure 4: 150th Eigenface

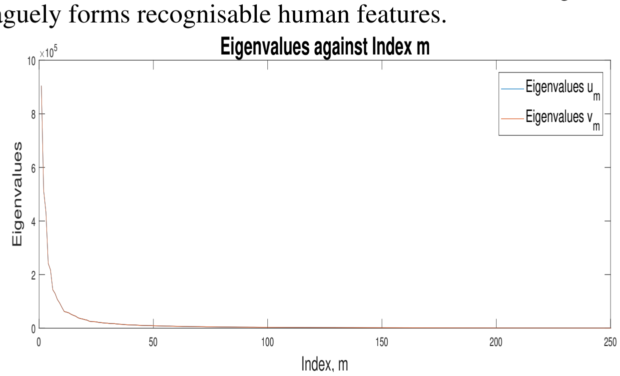

Figure 5: Variation of Eigenvalues

As $\boldsymbol{A}$ is a $D \times N$ matrix where $N \ll D$, although the dimension of $\boldsymbol{A}\boldsymbol{A}^T$ is $D \times D$, its maximum rank is only $N-1$ and thus the maximum number of non-zero eigenvalues possible is $N-1$. However, the rank may be lower than $N-1$ if the sample images are not linearly independent. Since no sample images were duplicated and edited in the training set, the sample images are linearly independent and an analysis of the covariance matrix $\boldsymbol{S}$ in MATLAB revealed 363 non-zero eigenvalues when the training set size was $N = 364$

3.2. Efficient Computation of Eigenfaces

As elaborated in Section 3.1 the matrix $\boldsymbol{S} = \frac{1}{N}\boldsymbol{A}\boldsymbol{A}^T$ has dimensions $D \times D$ and calculating its eigenvectors $\boldsymbol{u}_m$ is computationally expensive, especially for higher resolution, larger images. To improve computational time and efficiency for use in a real-time system, the eigenvectors of the smaller similar matrix $\boldsymbol{S}' = \frac{1}{N}\boldsymbol{A}^T\boldsymbol{A}$ can be computed instead. This is possible as the eigenvalues for similar matrices $\boldsymbol{S}'$ and $\boldsymbol{S}$ are the same and their eigenvectors are related by

$$\boldsymbol{u}_n = \frac{\boldsymbol{A}\boldsymbol{v}_n}{\|\boldsymbol{A}\boldsymbol{v}_n\|} \tag{3}$$

where, $\boldsymbol{v}_n$ is the n^{th} eigenvector of $\boldsymbol{S} = \frac{1}{N}\boldsymbol{A}^T\boldsymbol{A}$

It is important to note that the above is only faster and more efficient if $N \ll D$. If $N \approx D$, then this method may be less efficient since there is the additional calculation of $\boldsymbol{u}_n$ from $\boldsymbol{v}_n$. However, in face recognition, $N \ll D$ is often the case since the number of training samples is unlikely to be the same magnitude as the original subspace dimension. Table 1 below shows this time difference.

Method	$\boldsymbol{A}\boldsymbol{A}^T$	$\boldsymbol{A}^T\boldsymbol{A}$
Time Taken	6.145s	0.8682s

Table 1: Computational Time Difference

3.3. Application of Eigenfaces

The eigenfaces determined by PCA in Section 3.2 form a lower dimensional subspace from which reconstruction and classification of both training and test face images can be done. Although there are $N-1$ eigenfaces available, as shown in Figure 6, only the eigenfaces with larger corresponding eigenvalues contribute significantly and hence it would be prudent to determine the minimal number of eigenfaces M that would provide both acceptable reconstruction and classification performance while minimising the computational time needed for such operations.

3.3.1 Reconstruction Performance

As elaborated in Section 3.1, face images in both the training and testing set can be represented as a linear combination of the eigenfaces determined by PCA. This is done by projecting each normalised face image onto the chosen M dimensional 'face space', $\boldsymbol{w}_n$

$$\boldsymbol{w}_n = [a_{n1}, a_{n2}, ..., a_{nM}]^T \tag{4}$$

where $a_{nm} = \boldsymbol{\phi}_n^T\boldsymbol{u}_m$, $m = 1, ..., M$ and $\boldsymbol{\phi}_n$ is the normalised face.

Each face image can then be reconstructed by the following linear combination

$$\tilde{\boldsymbol{x}}_n = \bar{\boldsymbol{x}} + \sum_{m=1}^{M} a_{nm}\boldsymbol{u}_m \tag{5}$$

with the theoretical and training reconstruction error given by

$$J_{theoretical} = \sum_{m=M+1}^{D} \lambda_m \tag{6}$$

and the test reconstruction error given by

$$J_{test} = \frac{1}{N}\sum_{n=1}^{N}\|\boldsymbol{x}_n - \tilde{\boldsymbol{x}}_n\|^2 \tag{7}$$

By varying the number of eigenfaces M, the reconstruction error for both the training and test sets are shown in Figure below.

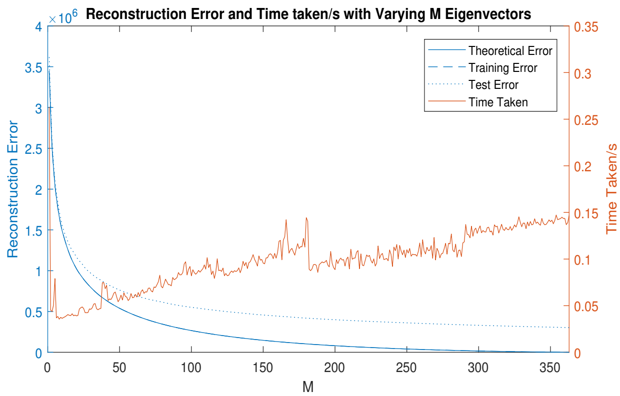

Figure 6: Variation of Reconstruction Error with M

Both the theoretical and training error are the same as shown in Equation 7 since they are calculated from the same eigenvalues.

3.3.2 Face Recognition Performance

This section explores two classifiers: 1) Classification by Nearest Neighbours and 2) Classification by Reconstruction (The Alternative Method). To start off, the Nearest Neighbour (NN) classification compares the projection $\boldsymbol{w}_n$of test images onto the computed face space of the training data and assigns the test image to the class with the minimum error.

As elaborated earlier in Section 3, face rec

3.4. Determining Optimal M Eigenfaces

4. Fisherfaces

Linear Discriminant Analysis (LDA) can be used as an alternative predictive model for our facial recognition problem. Unlike PCA, LDA is a discriminative model which extracts features that are useful for classification. LDA determines the directions that optimally separate data of different classes, uses these directions as a new subspace of dimension M_{lda}. In a c-class problem, LDA is used to determine M_{lda} discriminant functions $\mathbf{w}_i$ that would best classify the training data set.

In order to compute the discriminant functions, the between-class scatter matrix $\mathbf{S}_B$ and within-class scatter matrix $\mathbf{S}_W$ have to be calculated:

$$\mathbf{S}_B = \sum_{i=1}^{c}(\mathbf{m}_i - \mathbf{m})(\mathbf{m}_i - \mathbf{m})^\intercal$$

$$\mathbf{S}_W = \sum_{i=1}^{c}\sum_{x\in c_i}(\mathbf{x} - \mathbf{m}_i)(\mathbf{x} - \mathbf{m}_i)^\intercal$$

where $\mathbf{m}_i$ is the mean image of class c_i, and $\mathbf{m}$ is the mean image of the training set. The optimal projection $\mathbf{W}^*$ is chosen as the matrix with orthonormal columns which maximizes the ratios between the determinants of $\mathbf{S}_B$ and $\mathbf{S}_W$ of the projected samples, i.e.:

$$\mathbf{W}^* = \arg\max_{\mathbf{W}} \frac{|\mathbf{W}^\intercal\mathbf{S}_B\mathbf{W}|}{|\mathbf{W}^\intercal\mathbf{S}_W\mathbf{W}|} = \begin{pmatrix} | & | & & | \\ \mathbf{w}_1 & \mathbf{w}_1 & \dots & \mathbf{w}_{M_{lda}} \\ | & | & & | \end{pmatrix}$$

where $\mathbf{w}_i$ is the set of generalized eigenvectors of $\mathbf{S}_B$ and $\mathbf{S}_W$ corresponding to M_{lda} largest eigenvalues.

$$\mathbf{S}_W^{-1}\mathbf{S}_B\mathbf{w}_i = \lambda_i\mathbf{w}_i, i = 1, ..., M_{lda}$$

Note that there are at most $c-1$ nonzero generalized eigenvalues. In order to obtain the eigenvectors $\mathbf{S}_W$ has to be non-singular, which is difficult for the facial recognition problem as $rank\,(\mathbf{S}_W) \leq N - c << D$, and $\mathbf{S}_W \in \mathbb{R}^{D\times D}$.

4.1. PCA-LDA

In order to overcome this singularity problem, PCA is first used to project the training image set to a lower dimension space of dimension $M_{pca} \leq N - c$. LDA is then performed onto this reduced training set to further reduce the dimension to $M_{lda} \leq c - 1$.

The test image set is also projected onto this PCA-LDA subspace. Nearest neighbour classification was used to assign test images with the label of its closest training image neighbour via the euclidean distance metric.

This process was performed for a range of $\left(M_{pca}, M_{lda}\right)$ values, where $M_{lda} \leq M_{pca}$. The values $\left(M_{pca}, M_{lda}\right) = (130, 26)$ was found to give the highest test accuracy of 87.14%.

- INSERT 3D PLOT HERE
- elaborate how rank(Sw) and rank(Sb) generated by question2a.m is relevant
- confusion matrix
- example success and failure
- its 3am im tired and only started revising for computational optimisation im doomed theres so little math but the past year papers are so hard why did I sign up for this idk what to do gg help theres still HPC and HCR omg im ded three more days to exam then two days to HCR demo then one day to settle two PR reports and one HPC submission looks like I have to bring my laptop to the slopes already sigh such is life

5. Ensemble Learning

- TALK ABOUT HOW A BUNCH OF WEAK LEARNERS CAN AGGREGATE TO GENERATE DECENT RESULTS
- HOW THERE ARE DIFFERENT MODELS FOR ENSEMBLE LEARNING, PCA-SVM, PCA-LDA-SVM, PCA-LDA-NN, RANDOM FEATURE SPACE, RANDOM BAGGING

- WE USE MAJORITY VOTING, BECAUSE IF YOU TAKE THE AVERAGE, WTH DOES FACE 1.87123 EVEN MEAN?

5.1. Random Feature Selection

THIS IS SHIT AVERAGE PERFORMANCE

5.2. Random Sample Bagging

THIS IS SLIGHTLY BETTER, ONLY SLIGHTLY

5.3. The ruler

Here is stuff. this is an inlined equation: $x = 1$.
This is an equation as a separate chunk:

$$\begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix}$$

$$\mu\lambda_n = 10, \forall x \in \mathbf{R}^n$$

The LaTeX style defines a printed ruler which should be present in the version submitted for review. The ruler is provided in order that reviewers may comment on particular lines in the paper without circumlocution. If you are preparing a document using a non-LaTeX document preparation system, please arrange for an equivalent ruler to appear on the final output pages. The presence or absence of the ruler should not change the appearance of any other content on the page. The camera ready copy should not contain a ruler. (LaTeX users may uncomment the `\cvprfinalcopy` command in the document preamble.) Reviewers: note that the ruler measurements do not align well with lines in the paper — this turns out to be very difficult to do well when the paper contains many figures and equations, and, when done, looks ugly. Just use fractional references (e.g. this line is 095.5), although in most cases one would expect that the approximate location will be adequate.

5.4. Mathematics

Please number all of your sections and displayed equations. It is important for readers to be able to refer to any particular equation. Just because you didn't refer to it in the text doesn't mean some future reader might not need to refer to it. It is cumbersome to have to use circumlocutions like "the equation second from the top of page 3 column 1". (Note that the ruler will not be present in the final copy, so is not an alternative to equation numbers). All authors will benefit from reading Mermin's description of how to write mathematics: `http://www.pamitc.org/documents/mermin.pdf`.

5.5. Blind review

Blind review means that you do not use the words "my" or "our" when citing previous work. That is all. (But see below for techreports.)

Saying "this builds on the work of Lucy Smith [1]" does not say that you are Lucy Smith; it says that you are building on her work. If you are Smith and Jones, do not say "as we show in [7]", say "as Smith and Jones show in [7]" and at the end of the paper, include reference 7 as you would any other cited work.

An example of an acceptable paper:

> An analysis of the frobnicatable foo filter.
>
> In this paper we present a performance analysis of the paper of Smith *et al*. [1], and show it to be inferior to all previously known methods. Why the previous paper was accepted without this analysis is beyond me.
>
> [1] Smith, L and Jones, C. "The frobnicatable foo filter, a fundamental contribution to human knowledge". Nature 381(12), 1-213.

If you are making a submission to another conference at the same time, which covers similar or overlapping material, you may need to refer to that submission in order to explain the differences, just as you would if you had previously published related work. In such cases, include the anonymized parallel submission [**?**] as additional material and cite it as

> [1] Authors. "The frobnicatable foo filter", F&G 2014 Submission ID 324, Supplied as additional material `fg324.pdf`.

Finally, you may feel you need to tell the reader that more details can be found elsewhere, and refer them to a technical report. For conference submissions, the paper must stand on its own, and not *require* the reviewer to go to a techreport for further details. Thus, you may say in the body of the paper "further details may be found in [**?**]". Then submit the techreport as additional material. Again, you may not assume the reviewers will read this material.

Sometimes your paper is about a problem which you tested using a tool which is widely known to be restricted to a single institution. For example, let's say it's 1969, you have solved a key problem on the Apollo lander, and you believe that the CVPR70 audience would like to hear about your solution. The work is a development of your celebrated 1968 paper entitled "Zero-g frobnication: How being the only people in the world with access to the Apollo lander source code makes us a wow at parties", by Zeus *et al*.

You can handle this paper like any other. Don't write "We show how to improve our previous work [Anonymous, 1968]. This time we tested the algorithm on a lunar lander [name of lander removed for blind review]". That would be silly, and would immediately identify the authors. Instead write the following:

> We describe a system for zero-g frobnication. This system is new because it handles the following cases: A, B. Previous systems [Zeus et al. 1968] didn't handle case B properly. Ours handles it by including a foo term in the bar integral.
>
> ...
>
> The proposed system was integrated with the Apollo lunar lander, and went all the way to the moon, don't you know. It displayed the following behaviours which show how well we solved cases A and B: ...

Figure 7: Example of caption. It is set in Roman so that mathematics (always set in Roman: $B \sin A = A \sin B$) may be included without an ugly clash.

As you can see, the above text follows standard scientific convention, reads better than the first version, and does not explicitly name you as the authors. A reviewer might think it likely that the new paper was written by Zeus *et al*., but cannot make any decision based on that guess. He or she would have to be sure that no other authors could have been contracted to solve problem B.

FAQ: Are acknowledgements OK? No. Leave them for the final copy.

5.6. Miscellaneous

Compare the following:

`$conf_a$`	$conf_a$
`$\mathit{conf}_a$`	conf_a

See The TEXbook, p165.

The space after *e.g*., meaning "for example", should not be a sentence-ending space. So *e.g*. is correct, *e.g.* is not. The provided `\eg` macro takes care of this.

When citing a multi-author paper, you may save space by using "et alia", shortened to "*et al*." (not "*et. al.*" as "*et*" is a complete word.) However, use it only when there are three or more authors. Thus, the following is correct: " Frobnication has been trendy lately. It was introduced by Alpher [**?**], and subsequently developed by Alpher and Fotheringham-Smythe [1], and Alpher *et al*. [2]."

This is incorrect: "... subsequently developed by Alpher *et al*. [1] ..." because reference [1] has just two authors. If you use the `\etal` macro provided, then you need not worry about double periods when used at the end of a sentence as in Alpher *et al*.

For this citation style, keep multiple citations in numerical (not chronological) order, so prefer [1, **?**, **?**] to [**?**, 1, **?**].

6. Formatting your paper

All text must be in a two-column format. The total allowable width of the text area is $6\frac{7}{8}$ inches (17.5 cm) wide by $8\frac{7}{8}$ inches (22.54 cm) high. Columns are to be $3\frac{1}{4}$ inches (8.25 cm) wide, with a $\frac{5}{16}$ inch (0.8 cm) space between them. The main title (on the first page) should begin 1.0 inch (2.54 cm) from the top edge of the page. The second and following pages should begin 1.0 inch (2.54 cm) from the top edge. On all pages, the bottom margin should be 1-1/8 inches (2.86 cm) from the bottom edge of the page for 8.5×11-inch paper; for A4 paper, approximately 1-5/8 inches (4.13 cm) from the bottom edge of the page.

6.1. Margins and page numbering

All printed material, including text, illustrations, and charts, must be kept within a print area 6-7/8 inches (17.5 cm) wide by 8-7/8 inches (22.54 cm) high. Page numbers should be in footer with page numbers, centered and .75 inches from the bottom of the page and make it start at the correct page number rather than the 4321 in the example. To do this fine the line (around line 23)

```
%\ifcvprfinal\pagestyle{empty}\fi
\setcounter{page}{4321}
```

where the number 4321 is your assigned starting page.

Make sure the first page is numbered by commenting out the first page being empty on line 46

```
%\thispagestyle{empty}
```

6.2. Type-style and fonts

Wherever Times is specified, Times Roman may also be used. If neither is available on your word processor, please use the font closest in appearance to Times to which you have access.

MAIN TITLE. Center the title 1-3/8 inches (3.49 cm) from the top edge of the first page. The title should be in Times 14-point, boldface type. Capitalize the first letter of nouns, pronouns, verbs, adjectives, and adverbs; do not capitalize articles, coordinate conjunctions, or prepositions (unless the title begins with such a word). Leave two blank lines after the title.

AUTHOR NAME(s) and AFFILIATION(s) are to be centered beneath the title and printed in Times 12-point, non-boldface type. This information is to be followed by two blank lines.

The ABSTRACT and MAIN TEXT are to be in a two-column format.

MAIN TEXT. Type main text in 10-point Times, single-spaced. Do NOT use double-spacing. All paragraphs should be indented 1 pica (approx. 1/6 inch or 0.422 cm). Make sure your text is fully justified—that is, flush left and flush right. Please do not place any additional blank lines between paragraphs.

Figure and table captions should be 9-point Roman type as in Figures 7 and 8. Short captions should be centred.
Callouts should be 9-point Helvetica, non-boldface type. Initially capitalize only the first word of section titles and first-, second-, and third-order headings.

FIRST-ORDER HEADINGS. (For example, **1. Introduction**) should be Times 12-point boldface, initially capitalized, flush left, with one blank line before, and one blank line after.

SECOND-ORDER HEADINGS. (For example, **1.1. Database elements**) should be Times 11-point boldface, initially capitalized, flush left, with one blank line before, and one after. If you require a third-order heading (we discourage it), use

Figure 8: Example of a short caption, which should be centered.

Method	Frobnability
Theirs	Frumpy
Yours	Frobbly
Ours	Makes one’s heart Frob

Table 2: Results. Ours is better.

10-point Times, boldface, initially capitalized, flush left, preceded by one blank line, followed by a period and your text on the same line.

6.3. Footnotes

Please use footnotes[1] sparingly. Indeed, try to avoid footnotes altogether and include necessary peripheral observations in the text (within parentheses, if you prefer, as in this sentence). If you wish to use a footnote, place it at the bottom of the column on the page on which it is referenced. Use Times 8-point type, single-spaced.

6.4. References

List and number all bibliographical references in 9-point Times, single-spaced, at the end of your paper. When referenced in the text, enclose the citation number in square brackets, for example [**?**]. Where appropriate, include the name(s) of editors of referenced books.

6.5. Illustrations, graphs, and photographs

All graphics should be centered. Please ensure that any point you wish to make is resolvable in a printed copy of the paper. Resize fonts in figures to match the font in the body text, and choose line widths which render effectively in print. Many readers (and reviewers), even of an electronic copy, will choose to print your paper in order to read it. You cannot insist that they do otherwise, and therefore must not assume that they can zoom in to see tiny details on a graphic.

When placing figures in LaTeX, it’s almost always best to use `\includegraphics`, and to specify the figure width as a multiple of the line width as in the example below

```
\usepackage[dvips]{graphicx} ...
\includegraphics[width=0.8\linewidth]
                {myfile.eps}
```

6.6. Color

Please refer to the author guidelines on the CVPR 2016 web page for a discussion of the use of color in your document.

7. Final copy

You must include your signed IEEE copyright release form when you submit your finished paper. We MUST have this form before your paper can be published in the proceedings.

References

[1] A. Alpher and J. P. N. Fotheringham-Smythe. Frobnication revisited. *Journal of Foo*, 13(1):234–778, 2017.

[2] A. Alpher, J. P. N. Fotheringham-Smythe, and G. Gamow. Can a machine frobnicate? *Journal of Foo*, 14(1):234–778, 2004.

[3] K. J. Law enforcement’s use of facial recognition technology.

[4] M. A. Turk and A. P. Pentland. Face recognition using eigenfaces. 1991.

[5] K. G. Waldemar Wjcik and M. Junisbekov. Face recognition - semisupervised classification, subspace projection and evaluation methods. 2016.

[1]This is what a footnote looks like. It often distracts the reader from the main flow of the argument.