Towards a Visual Human Recognition using a Monocular Camera on a Drone

By

Andrews Frimpong Adu (andrews@aims.edu.gh)

June 2016

AN ESSAY PRESENTED TO AIMS-GHANA IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE AWARD OF MASTER OF SCIENCE IN MATHEMATICS

AIMS | African Institute for Mathematical Sciences GHANA

DECLARATION

This work was carried out at AIMS-Ghana in partial fulfilment of the requirements for a Master of Science Degree.

I hereby declare that except where due acknowledgement is made, this work has never been presented wholly or in part for the award of a degree at AIMS-Ghana or any other University.

Scan your signature

Student: Firstname Middlename Surname

Scan your signature

Supervisor: Firstname Middlename Surname

ACKNOWLEDGEMENTS

This is optional and should be at most half a page. Thanks Ma, Thanks Pa. One paragraph in normal language is the most respectful.

Do not use too much bold, any figures, or sign at the bottom.

DEDICATION

This is optional.

Abstract

A short, abstracted description of your essay goes here. It should be about 100 words long. But write it last.

An abstract is not a summary of your essay: it's an abstraction of that. It tells the readers why they should be interested in your essay but summarises all they need to know if they read no further.

The writing style used in an abstract is like the style used in the rest of your essay: concise, clear and direct. In the rest of the essay, however, you will introduce and use technical terms. In the abstract you should avoid them in order to make the result comprehensible to all.

You may like to repeat the abstract in your mother tongue.

Contents

1. Introduction

This is a textual citation Shannon (1993). And this is a parenthetical citation (Shannon, 1993). You probably want to use the latter more often.

1.1 Moving On

Let's demonstrate a figure by looking at Fig. 1.1.

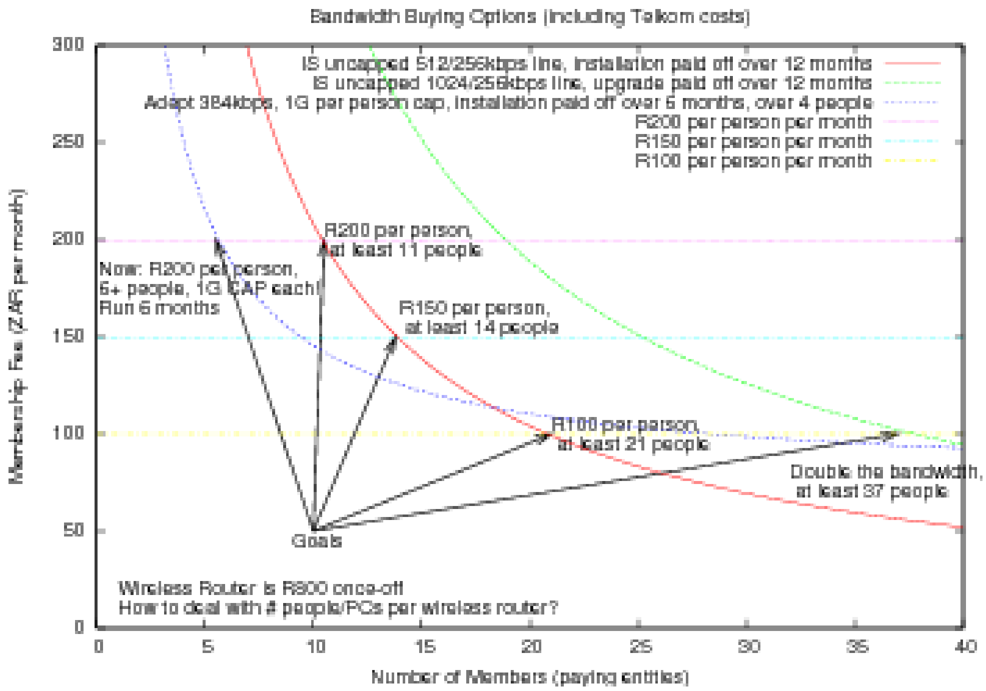

Figure 1.1: Planning community bandwidth sharing costs. Note caption capitalization.

Remember how to include code with `verbatim` and to fix the tabs in python in a verbatim environment? It may be best to have an 'include' command for code, not to have to re-edit it all the time.

```
# This program prints hello

import Scipy as S

if __name__ == "__main__":
        print "hello"
```

2. Related Work

2.0.1 Theorem (Jeff's Washing Theorem)**.** *If an item of clothing is too big, then washing it makes it bigger; but if it is too small, washing it makes it smaller.*

Proof. Stated without proof. But a proof would look like this. □

Notice that no Lemmas are required in the proof of Theorem 2.0.1.

3. Face Recognition

In this chapter, three approaches to the face recognition process are described, namely; Eigenfaces method for face recognition, Fisherfaces method and Local binary patterns (LBP).

The Eigenfaces method for face recognition is based on Principal Component Analysis (PCA), which linearly projects a high-dimensional image space into a low-dimensional feature space called face space aiding efficient feature classification. The use of eigenfaces for face recognition was pioneered by Kirby and Sirovich (Sirovich and Kirby, 1987) and used by Turk and Pentland (Turk and Pentland, 1991) in automated face recognition systems.

The Eigenfaces method is known by many as the first successful facial recognition technology which has served as the basis for one of the top commercial face recognition technology products. There have been many extensions and many new developments in automatic recognition systems using the Eigenfaces method since its initial development. The name Eigenfaces is given to the set of eigenvectors obtained from the covariance matrix of the probability distribution over the high-dimensional vector space of face images.

The Fisherfaces method, proposed by Belhumeur, Hespanha and Kriegman, (Belhumeur et al., 1997) is based on Fisher's linear discriminant (FLD) for face recognition. The Fisherfaces method uses discriminant analysis to project the initial high-dimensional image space into the directions that best separate the classes, by maximizing the ratio of between-class scatter to that of within-class scatter.

Local Binary Pattern (LBP) is a texture description technique proposed originally for texture analysis. It was introduced in 1996 and 2002 (Ojala et al., 2002) based on the assumption that texture has locally two complementary aspects, a pattern and its strength. In image analysis, LBP is used to determine the local features in facial images. The facial image is divided into local regions from which LBP histograms (LBP features) are formed and concatenated into a single feature vector (histogram) which efficiently represent the face image. The feature vector is then used to measure the similarities by calculating the distance between the images.

3.1 Eigenfaces

Eigenface recognition technique is based on the use of principal component analysis (PCA). Principal component analysis, (PCA), is a linear transformation technique often used for feature extraction and data dimension reduction problems. PCA was proposed by Karl Pearson (1901) and Harold Hotelling (1933). Its goal is to extract important features from the image and to express this features as a set of new orthogonal variables called principal components (Turk and Pentland, 1991). Thus, given a set $\mathbf{X}$ consisting of M images, PCA produces a new set of images known as feature space, and the elements in these space are uncorrelated and are ordered in terms of the amount of variance they explain from the original image set. The important features extracted are known as "Eigenfaces" because they are the eigenvectors of the image set.

3.1.1 Calculating Eigenfaces (Principal components). Let $\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_M\}$ be a set of M sample face images formed by the vectors $\mathbf{x_k} \in \mathbb{R}^N$. The average $\bar{\mathbf{x}}$ (mean face) and the covariance matrix $\mathrm{Cov}(\mathbf{x}) = \mathbf{\Sigma}$ of these image set are defined by:

$$\bar{\mathbf{x}} = \frac{1}{M}\sum_{n=1}^{M} \mathbf{x_n}$$

and

$$\mathbf{\Sigma} = \frac{1}{M}\sum_{n=1}^{M} (\mathbf{x_n} - \bar{\mathbf{x}})(\mathbf{x_n} - \bar{\mathbf{x}})^T$$

respectively. Then, each face image differs from the mean face by the vector $\phi_i = \mathbf{x}_i - \bar{\mathbf{x}}$, where $i = 1, 2, \ldots, M$. PCA seeks a linear transformation that maps the original $N-$dimensional image space into a $d-$dimensional feature space, where $d < N$, with minimum loss of information. Thus, PCA seek vectors, $\mathbf{w}$, such that the sample, after projection onto $\mathbf{w}$, is most spread out. The projection of all the data points onto this dimension can be written as:

$$\mathbf{z} = \mathbf{w}^T \mathbf{x}_n$$

and the variance of the projected data is given by:

$$\mathrm{var}(\mathbf{z}) = \mathbf{w}^T \mathbf{\Sigma} \mathbf{w}. \tag{3.1.1}$$

We seek $\mathbf{w}$ that maximized $\mathrm{var}(\mathbf{z})$ subject to the constraint that $\mathbf{w}^T\mathbf{w} = 1$ (since $||\mathrm{w}|| \longrightarrow \infty$). Writing this as optimization problem, we have

$$\begin{aligned} \max\ & \mathbf{w}^T \mathbf{\Sigma} \mathbf{w} \\ \text{subject to } & \mathbf{w}^T \mathbf{w} = 1. \end{aligned} \tag{3.1.2}$$

Introducing the Lagrangian multiplier, denoted by λ, and writing $(3.1.2)$ as Lagrange problem, we have

$$L(\mathbf{w}, \lambda) = \mathbf{w}^T \mathbf{\Sigma} \mathbf{w} - \lambda(\mathbf{w}^T \mathbf{w} - 1).$$

Taking the derivative with respect to $\mathbf{w}$ and setting it equal to 0, we have

$$\begin{aligned} 2\mathbf{\Sigma}\mathbf{w} - 2\lambda\mathbf{w} &= 0 \\ 2\mathbf{\Sigma}\mathbf{w} &= 2\lambda\mathbf{w} \\ \Longrightarrow \mathbf{\Sigma}\mathbf{w} &= \lambda\mathbf{w} \end{aligned}$$

where $\mathbf{w}$ is an eigenvector of $\mathbf{\Sigma}$ and λ is the corresponding eigenvalue. Pre-multiplying both sides by $\mathbf{w}^T$ we have

$$\mathbf{w}^T \mathbf{\Sigma} \mathbf{w} = \lambda \mathbf{w}^T \mathbf{w} = \lambda$$

and so the variance will be a maximum when we set $\mathbf{w}$ equal to the eigenvector with the highest eigenvalue λ. Therefore the first principal component is given by the eigenvector with the largest associated eigenvalue of the sample covariance matrix $\mathbf{\Sigma}$. A similar argument can show that the k dominant eigenvectors of covariance matrix $\mathbf{\Sigma}$ determine the first k principal components. In general

$$\mathrm{z_k} = \mathbf{W}^T \mathbf{\Sigma} \mathbf{W} = \lambda_k$$

where $\mathbf{W} = [\mathbf{w}_1, \mathbf{w}_2, \cdots, \mathbf{w}_k]$.

3.1.2 Recognition. The Eigenfaces forms a basis for the new dimensional feature space (subspace) and spans the original data set. The images in the original data are linear combinations of the feature vectors (Eigenfaces / weights vector). A new input data forms a linear combination of these weight vectors and compare distances between the images in the training set. The closer the images, the more likely it recognizes as a facial image.

The projection operation characterizes an individual face by a weighted sum of the eigenface features, and so to recognize a particular face it is necessary only to compare these weights to those of known individuals.

3.2 Fisherfaces

Fisherfaces technique technique is a feature extraction technique based on Fisher's linear discriminant analysis (FLDA). It seeks to find a projection, $\mathbf{w}$, that maximizes the between-class separability and minimizes the within-class variability. The technique first project the face images from the original vector space to a lower dimensional space (face subspace) using Principal Component Analysis (PCA) and then LDA is applied next to find the best linear discriminant features on that face subspace. Recognition is then done by using nearest neighbor (NN) classifier in this final subspace.

3.2.1 Calculating Fisherfaces. Consider two classes with data points n_1 and n_2. Let $\{\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_M\}$, $\mathbf{x} \in \mathbb{R}^N$ be the set of M face images belonging to the two classes, C_1 and C_2. FLD seeks a direction, defined by the vector $\mathbf{w}$, such that when the data are projected onto $\mathbf{w}$, the data points from the two classes are as well separated as possible. We define

$$\mathbf{z} = \mathbf{w}^T \mathbf{x} = \sum_{k=1}^{M} \mathbf{w}_k \mathbf{x}_k$$

to be the projection of $\mathbf{x}$ onto $\mathbf{w}$ and thus a dimensionality reduction from N to 1. The means for the two classes are defined as:

$$\mu_1 = \frac{1}{n_1} \sum_{k \in C_1} \mathbf{x}_k \quad \text{and} \quad \mu_2 = \frac{1}{n_2} \sum_{k \in C_2} \mathbf{x}_k$$

and the means after projection are given by

$$\mathbf{w}^T \mu_1 \quad \text{and} \quad \mathbf{w}^T \mu_2$$

respectively. The covariance matrix after projection are:

$$\mathbf{\Sigma}_1 = \mathbf{w}^T \mathbf{\Sigma}_1 \mathbf{w}$$

and

$$\mathbf{\Sigma}_2 = \mathbf{w}^T \mathbf{\Sigma}_2 \mathbf{w}.$$

For the two classes to be as well separated as possible after projection, we would like the means to be as far as possible and the data points of the classes be scattered in as small a region as possible. Thus, the distance between the projected class means is:

$$\begin{aligned}\left(\mathbf{w}^T\mu_1 - \mathbf{w}^T\mu_2\right)^2 &= \left(\mathbf{w}^T\mu_1 - \mathbf{w}^T\mu_2\right)^T \left(\mathbf{w}^T\mu_1 - \mathbf{w}^T\mu_2\right) \\ &= \left(\mu_1 - \mu_2\right)\mathbf{w}\mathbf{w}^T\left(\mu_1 - \mu_2\right) \\ &= \mathbf{w}^T\left(\mu_1 - \mu_2\right)\left(\mu_1 - \mu_2\right)\mathbf{w} \\ &= \mathbf{w}^T S_b \mathbf{w}\end{aligned}$$

where $S_b = (\mu_1 - \mu_2)(\mu_1 - \mu_2)$ is the between class variance. Since we want to minimize the variance within each class, we write

$$\begin{aligned}\mathbf{w}^T\mathbf{\Sigma}_1\mathbf{w} + \mathbf{w}^T\mathbf{\Sigma}_2\mathbf{w} &= \mathbf{w}^T\left(\mathbf{\Sigma}_1 + \mathbf{\Sigma}_2\right)\mathbf{w} \\ &= \mathbf{w}^T S_w \mathbf{w}\end{aligned}$$

where $S_w = \mathbf{\Sigma}_1 + \mathbf{\Sigma}_2$ is the within class covariance. Fisher's Linear discriminant is $\mathbf{w}$ such that the distance between projected class means are maximize and the within class variance are minimize by maximizing the ratio:

$$\max_{\mathbf{w}} \frac{\mathbf{w}^T S_b \mathbf{w}}{\mathbf{w}^T S_w \mathbf{w}}$$

and this is equivalent to finding :

$$\begin{aligned}&\max_{\mathbf{w}} \mathbf{w}^T S_b \mathbf{w} \\ &\text{ s.t. } \mathbf{w}^T S_w \mathbf{w} = 1.\end{aligned}$$

Applying Lagrange multiplier, denoted by λ, we have

$$L(\mathbf{w}, \lambda) = \mathbf{w}^T S_b \mathbf{w} - \lambda\left(\mathbf{w}^T S_w \mathbf{w} - 1\right).$$

Taking the derivative with respect to $\mathbf{w}$ and setting to zero, we get:

$$\begin{aligned}2S_b\mathbf{w} - 2\lambda S_w \mathbf{w} &= 0 \\ S_b\mathbf{w} &= \lambda S_w \mathbf{w}\end{aligned}$$

which is the generalized eigenvalue problem that is equivalent to (for S_w non-singular):

$$S_w^{-1} S_b \mathbf{w} = \lambda \mathbf{w}$$

where λ and $\mathbf{w}$ are eigenvalues and eigenvectors of $S_w^{-1}S_b$ respectively. $\mathbf{w}$ is the eigenvector corresponding to the largest eigenvalue of $S_w^{-1}S_b$.

3.2.2 Recognition.

3.3 Local Binary Patterns (LBP)

In the local binary patterns (LPB) technique for face description, the face image is divided into small blocks called local regions, from which LBP descriptors are extracted into histograms. The obtained histograms from the local regions are concatenated into a single feature histogram, which gives a global description of the face image (Figure 3.1). During image recognition, images are compared by measuring the similarity (distance) between their histograms.

The method considers a 3×3 block of pixels and then LBP is derived by comparing the center pixel with its neighbors to derive a code which is stored at the center pixel. A neighboring pixel is assigned the binary number 1 if its intensity is greater than or equal to the center pixel and 0 otherwise. This gives a binary pattern of eight digit with $2^8 = 256$ possible combinations called Local Binary Patterns or LBP codes.

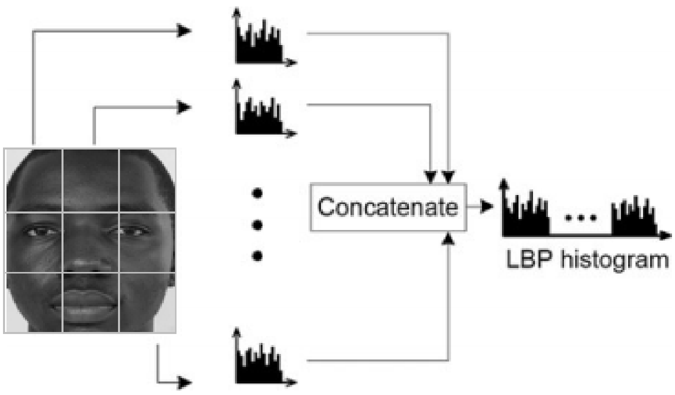

Figure 3.1: Face image divided into 9 regions with every region representing a histogram and then a concatenated histogram.

3.3.1 Calculating LBP. For the central pixel P and the neighboring pixels $\mathbf{P}_x$, the process depends on thresholding, which is the function

$$\rho(x) = \begin{cases} 1, & \mathbf{P}_x \geq P \\ 0, & \text{Otherwise.} \end{cases} \tag{3.3.1}$$

The LBP code is derived as a result of binary weighting applied to the result of thresholding which is equivalent to thresholding the points neighboring the center point and then unwrapping the code as a binary code. The LBP code for a point P with eight neighbors x is then given by

$$\mathrm{LBP} = \sum_{x \in 1,8} \rho(x) \times 2^{x-1}. \tag{3.3.2}$$

$$\tag{3.3.3}$$

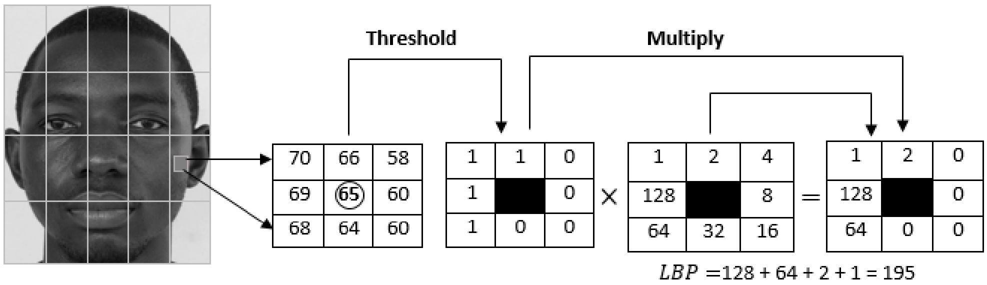

Figure 3.2: The original LBP operator.

From Fig. 3.2, the binary code and the LBP code are calculated from the weights (2^p) as follows:

$$\begin{aligned}
\text{Binary code} &: 10110010 \\
\text{LBP code} &: 1 \times 2^7 + 1 \times 2^6 + 0 \times 2^5 + 0 \times 2^4 + 0 \times 2^3 + 0 \times 2^2 + 1 \times 2^1 + 1 \times 2^0 \\
&= 128 + 64 + 2 + 1 \\
&= 195.
\end{aligned}$$

3.3.2 Recognition. Face images are identified using k-nearest neighbor classifier by majority vote of its neighbors. Consider a training set $X = \{\mathbf{I}_1, \mathbf{I}_2, \ldots, \mathbf{I}_m\}$ of m different people with each $\mathbf{I}$ having n images. Given an input image Y that is to be identified, the distances among all the $n = n_1 + n_2 + \cdots + n_m$ images and Y are computed. Then, the person of image Y is identified as the person with the most images in the k-nearest neighbors. The distance between the histograms are computed by the use of Euclidean distance. The histogram distance between the histogram H_i of image Y which is to be identified and the histogram H_o of the image O which is given in the training set is calculated as:

$$D_E(Y, O) = \sqrt{\sum_{k=2}^{l} (H_i(k) - H_d(k))^2}.$$

4. Face Tracking

4.1 Procedure

5. Experiments

References

Alan Adolphson, Steven Sperber, and Marvin Tretkoff, editors. *p-adic Methods in Number Theory and Algebraic Geometry*. Number 133 in Contemporary Mathematics. American Mathematical Society, Providence, RI, 1992.

Alan Beardon. From problem solving to research, 2006. Unpublished manuscript.

Peter N Belhumeur, João P Hespanha, and David J Kriegman. Eigenfaces vs. fisherfaces: Recognition using class specific linear projection. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 19(7):711–720, 1997.

Matthew Davey. *Error-correction using Low-Density Parity-Check Codes*. Phd, University of Cambridge, 1999.

Anil K Jain and Stan Z Li. *Handbook of face recognition*, volume 1. Springer, 2005.

Leslie Lamport. *LaTeX: A Document Preparation System*. Addison-Wesley, 1986.

D. J. C. MacKay and R. M. Neal. Good codes based on very sparse matrices. Available from www.inference.phy.cam.ac.uk, 1995.

David MacKay. Statistical testing of high precision digitisers. Technical Report 3971, Royal Signals and Radar Establishment, Malvern, Worcester. WR14 3PS, 1986a.

David MacKay. A free energy minimization framework for inference problems in modulo 2 arithmetic. In B. Preneel, editor, *Fast Software Encryption (Proceedings of 1994 K.U. Leuven Workshop on Cryptographic Algorithms)*, number 1008 in Lecture Notes in Computer Science Series, pages 179–195. Springer, 1995b.

Timo Ojala, Matti Pietikäinen, and Topi Mäenpää. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 24(7):971–987, 2002.

Nicolai Petkov and Michel A Westenberg. *Computer Analysis of Images and Patterns: 10th International Conference, CAIP 2003, Groningen, The Netherlands, August 25-27, 2003, Proceedings*, volume 2756. Springer, 2003.

Matti Pietikäinen. Local binary patterns. *Scholarpedia*, 5(3):9775, 2010.

Claude Shannon. A mathematical theory of communication. *Bell Sys. Tech. J.*, 27:379–423, 623–656, 1948.

Claude Shannon. The best detection of pulses. In N. J. A. Sloane and A. D. Wyner, editors, *Collected Papers of Claude Shannon*, pages 148–150. IEEE Press, New York, 1993.

Lawrence Sirovich and Michael Kirby. Low-dimensional procedure for the characterization of human faces. *Josa a*, 4(3):519–524, 1987.

Matthew A Turk and Alex P Pentland. Face recognition using eigenfaces. In *Computer Vision and Pattern Recognition, 1991. Proceedings CVPR'91., IEEE Computer Society Conference on*, pages 586–591. IEEE, 1991.

Web12. Commercial mobile robot simulation software. Webots, www.cyberbotics.com, Accessed April 2013.

Wik12. Black scholes. Wikipedia, the Free Encyclopedia, http://en.wikipedia.org/wiki/Black%E2%80%93Scholes, Accessed April 2012.