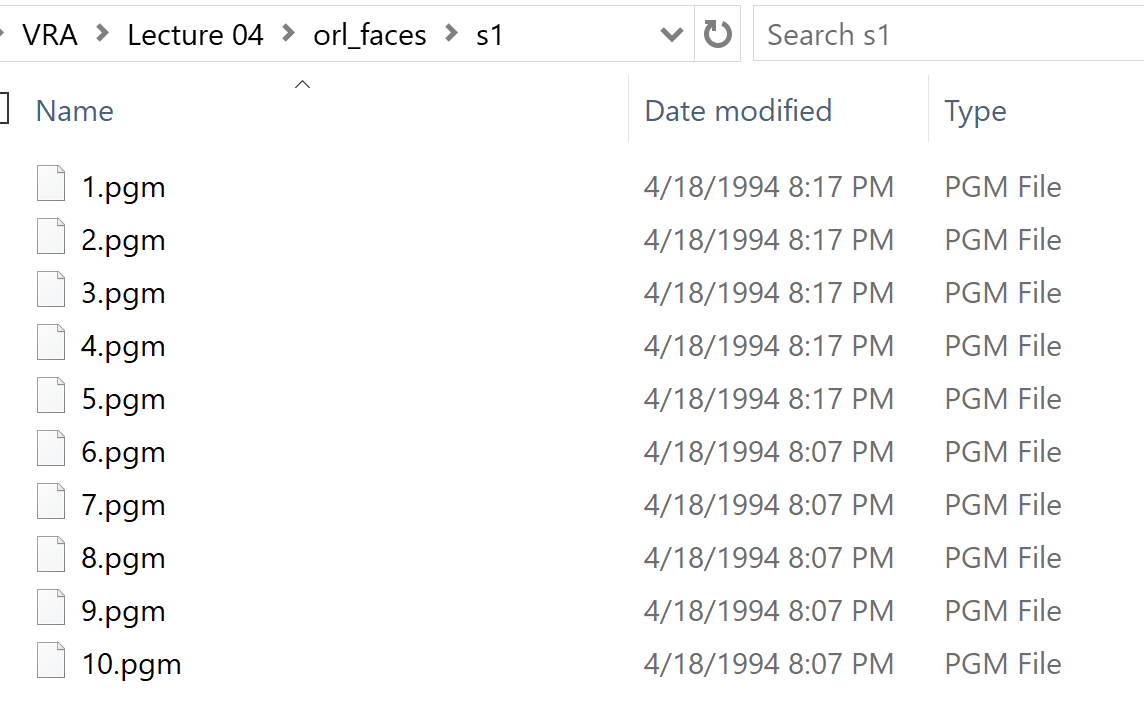
# Báo cáo tìm hiểu về ‘The Database of Face’

[*http://www.cl.cam.ac.uk/research/dtg/attarchive/facedatabase.html*](http://www.cl.cam.ac.uk/research/dtg/attarchive/facedatabase.html)

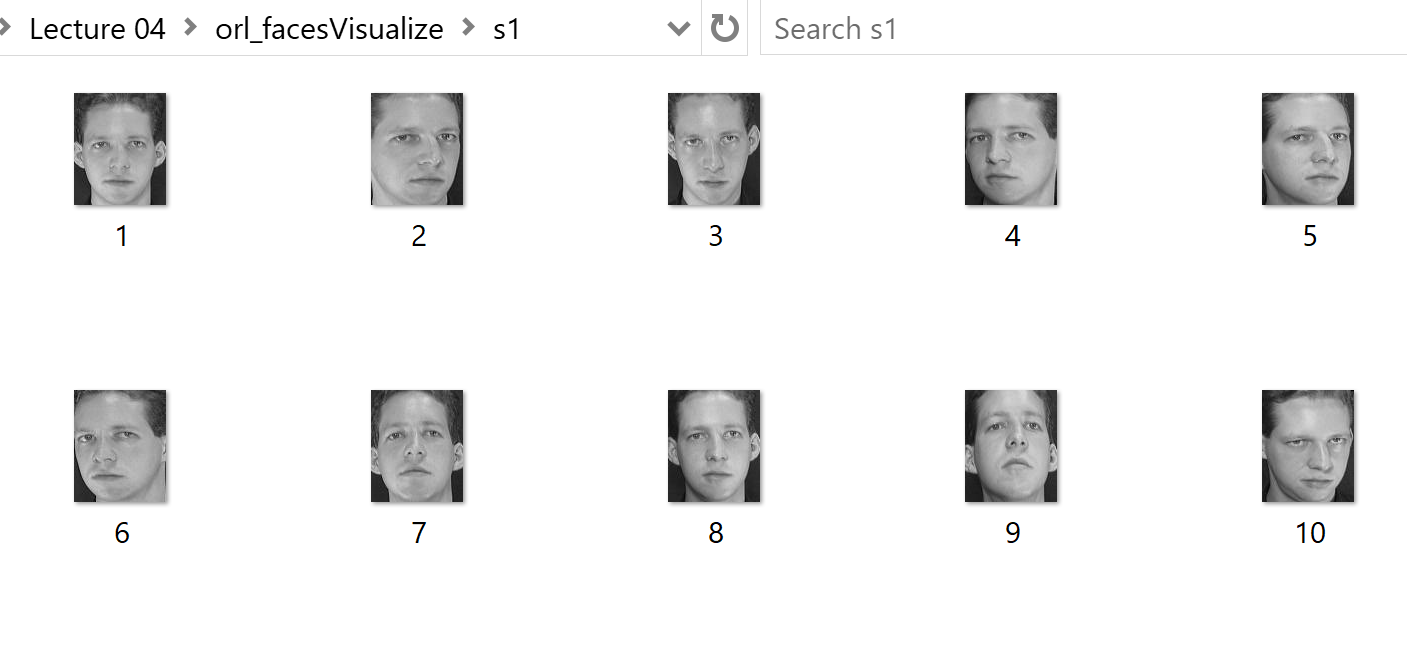
## Tổ chức thư mục

Dữ liệu tải về gồm 1 tập tin **att\_faces.zip**, sau khi giải nén được một thư mục ‘orl\_faces’. Thư mục ‘orl\_faces’ gồm 40 thư mục con lần lượt là s1, s2, …, s40. Mỗi thư mục con chứ 10 tập tin x.pgm (x nhận giá trị từ 1 đến 10.)

40 thư mục con là dữ liệu của 40 người. Mỗi người gồm có 10 tập tin pgm.



Hình 1 – Các tập tin pgm trong thư mục ‘s1’



Hình 2 – Kết quả Visualize các tập tin pgm trong thư mục ‘s1’

## Tổ chức dữ liệu huấn luyện và kiểm tra

Để áp dụng bài tập 3, cần chuẩn bị dữ liệu gần giống chữ số viết tay. Mà bộ dữ liệu ‘The Database of Face’ chỉ có 1 thư mục ‘orl\_faces’, không có phân biệt danh sách huấn luyện và kiểm tra.

Bộ dữ liệu được tổ chức lại như sau:

* Thư mục DataFaceTrain: gồm 40 thư mục con (s1, s2,…, s40). Mỗi thư mục chứ 7 ảnh huấn luyện. Số lượng ảnh huấn luyện là 280 ảnh.
* Thư mục DataFaceTest: gồm 40 thư mục con (s1, s2,…, s40). Mỗi thư mục chứ 3 ảnh kiểm tra. Số lượng ảnh kiểm tra là 120 ảnh.

Các ảnh cùng thư mục con s1 sẽ có nhãn là 1. Các ảnh còn lại tương tự.

Mỗi ảnh có kích thước là 112 x 96, sẽ được chuyển thành 1 vecto cột 10304 chiều.

Tất cả các ảnh huấn luyện sẽ được lưu vào tập tin **imgTrainImagesAll.mat** là ma trận 10304 x 280. Nhãn tương ứng cho các ảnh được lưu vào tập tin **lblTrainLabelsAll.mat**  là ma trận 1x 280.

Tất cả các ảnh huấn luyện sẽ được lưu vào tập tin **imgTestImagesAll.mat** là ma trận 10304 x 120. Nhãn tương ứng cho các ảnh được lưu vào tập tin **lblTestLabelsAll.mat**  là ma trận 1 x120.

Sau khi tổ chức lại, chúng ta thu được 4 tập tin biễu diễu dữ liệu là: **imgTrainImagesAll.mat, lblTrainLabelsAll.mat**, **imgTestImagesAll.mat, lblTestLabelsAll.mat.**