Subject:
Year: Month: Date:

PCA : روشی برای کاهش ابعاد است.

(۱) بردارهای ویژه: یکه و متعامد هستند و ما هر برداری را می‌توانیم به صورت ترکیبی از این بردارهای ویژه بنویسیم.

مقادیر ویژه متناظر با

E یک ماتریس قطری است که بردارهای ویژه روی قطر آن هستند.

$$E=\begin{bmatrix}\lambda_1 & & 0\\ \vdots & \ddots & \vdots\\ & \cdots & \lambda_n\end{bmatrix}$$

V هم ماتریس شامل بردارهای ویژه U می‌باشد.

$$V=\begin{bmatrix}u_1\\ \vdots\\ u_n\end{bmatrix}$$

u : بردار ویژه   λ = مقدار ویژه

در جهت این بردارهای ویژه ما به اندازه $\sqrt{\lambda}$ مقدار داریم و اگر این λ ها از بزرگ به کوچک مرتب کنیم در جهت بعضی از آنها مقادیر کمی داریم پس می‌توانیم یک تعدادی از این بردارهای ویژه که در راستای آنها مقادیر کمی را داریم حذف کنیم. اینکه بدانیم از کجا کم کنیم؟ یک مؤلفه‌ای به نام واریانس تعریف می‌کنند که هر چه واریانس مقادیر بیشتر نگه داشته بشود، دیتا بیشتر نگه داشته می‌شود.

k تا نگه داشتیم

$$V_{\text{واریانس}} = \frac{\sum_{i=1}^{k}\lambda_i}{\sum_{i=1}^{n}\lambda_i \ (\text{کل})}$$

مثلاً اگر V=99% باشد تا آن k پیش می‌رویم که 99% اطلاعات را برای ما نگه دارد.

به دو روش می‌توانیم V و E را به دست آوریم: svd / eig value. در اینجا از eig استفاده می‌کنیم که مقادیر ویژه و بردار آنها را به ما می‌دهد.

در اینجا ما یک x داریم که ابعاد آن خیلی زیاد است، می‌خواهیم ابعاد آن را کم کنیم.

$$x_{\text{جدید}} = x_{rot} = U^T x_{\text{قبلی}}$$

x_{rot} همان ابعاد x قبلی را دارد. V , λ را نیز داریم.

در x_{rot}، ابتدا بردارهای ویژه (که ممکن است در x قبلی مرتب نباشند) را مرتب می‌کنیم و متناظر با آن بردارهای ویژه، U ها را می‌نویسیم. حال x قبلی را در این بردارهای U مرتب شده ضرب می‌کنیم. x جدید همان اطلاعات قبلی را دارد فقط داده را انتقال داده‌ایم چون پراکندگی داده‌ها بیشتر در جهت بردارهای ویژه است. بردارهای ویژه را از ↑ به ↓ مرتب کردیم و حالا برای حذف کردن از پایین حذف می‌کنیم. یعنی از x_{rot} اگر بخواهیم k تا را نگه داریم مؤلفه‌های k+1 تا آخر را صفر می‌کنیم.

اگر U را در x_{rot} ضرب کنیم همان x قبلی را می‌سازیم:

$$x_{rot} = \underset{\text{U را داریم}}{U^T x} \Longrightarrow U x_{rot} = \underbrace{UU^T}_{I} x = x$$

PAPCO

Subject:
Year: Month Date:

پس از x_{rot} های اطلاعات x را داریم و خوبی آن این است که می‌دانیم از کجا اطلاعات کم‌ارزش‌تر را حذف کنیم.

سفیدکردن:

بدین معناست که در حقیقت همه بردارها دیگر نسبت به هم تفاوت سطوح نداریم و در واقع جهت پراکندگی بیشتری نداریم یا یکسان. برای همین بردارها را تقسیم بر لانداها می‌کنیم.
هر x_{rot} هر کدام از مؤلفه‌هایش به اندازه λ مقدار دارد. اگر بخواهیم نسبت این را سفید کنیم (یعنی ما در همه مؤلفه‌های مؤلفه‌شان یک مقدار داشته باشیم و عوارض سیگنال آن را بکشیم یک حفظ شود) کافی است $\times \frac{1}{\sqrt{\lambda}}$ بکنیم. ماتریس جدید را اگر بردارهایش را بدست بیاوریم در اطراف هر بردار دیگر یک مقدار دارد. اما ممکن است لاندا صفر شود و برای اینکه بتوانیم $\frac{1}{\sqrt{\lambda}}$ را انجام دهیم با یک ε جمع می‌کنیم.
اگر چه هر چه ε را ؟ کنیم از سفیدکردن فاصله می‌گیریم اما دیگر مشکل تقسیم بر ۰ را نداریم.

توضیح قسمت‌های کد:

steps:

$X_{144\times10000}$ ما باید ها تنها ۱۴۴ برای ما مهم است لذا میانگین ۱۴۴ را حول ۰ حساب می‌کنیم.

mean2x = mean(x,2)

$x = x - \text{repmat}(\text{mean2x}, [1, \underbrace{\text{size}(x,2)}_{10000}])$

برای اینکه بتوانیم از x کم کنیم باید ماتریس‌ها را هم اندازه کنیم. ماتریس mean را به اندازه [1,10000] تکرار می‌کند.

محاسبه xrot:

$covx = x.x^T / 10000 \rightarrow$ حساب می‌کنیم.

$[U,E] = eig(covx) \longrightarrow$ بردارهای V,E را بدست می‌آوریم.

$E = diag(E)$ دستور diag ماتریس قطری را تبدیل به بردار ستونی می‌کند.

ایندکس

$[B,I] = sort(-1 * E]$ حال با استفاده از دستور sort مرتب می‌کنیم.

با sort از ... به ... را می‌خواهیم درحالیکه sort عکس است پس در ۱- ضرب می‌کنیم

$U = U(:,I) \longrightarrow$ مرتب‌سازی U

$\Rightarrow x_{Rot} = U^T . X$

PAPCO

Subject:
Year Month Date

Step2:

محاسبه k : k را به ازای حفظ درصد خواسته شده بدست می‌آوریم.

$V = 90\% \Rightarrow k = 15$ (15/144)

Step3:

$V = 99\% \Rightarrow k = 89$ (89/144)

حالا که بعد از بدست آوردن k حال می‌خواهیم ابعاد را کاهش دهیم. برای اینکار در x_{rot} از سطر k+1 تا آخر، همه سطرهای ماتریس را 0 می‌کنیم.

$reduced x_{rot}(k+1 : end, :) = 0$

Step 4a

بدست آوردن x اولیه:

هرچه ε به 0 نزدیکتر باشد شبیه‌تر ... $\frac{1}{\sqrt{\lambda+\varepsilon}}$ به 1 نزدیکتر می‌شود. که اگر λ ها نزدیک به صفر باشند این تقسیم نزدیک 1 اما در عمل λ ها نزدیک 0 هستند.

ε روی سفید کردن اثر می‌گذارد.

$$x_{PCAwhite} = \underbrace{x_{rot}}_{U^T \cdot x} \times diag\left(\frac{1}{\sqrt{E+\varepsilon}}\right)$$

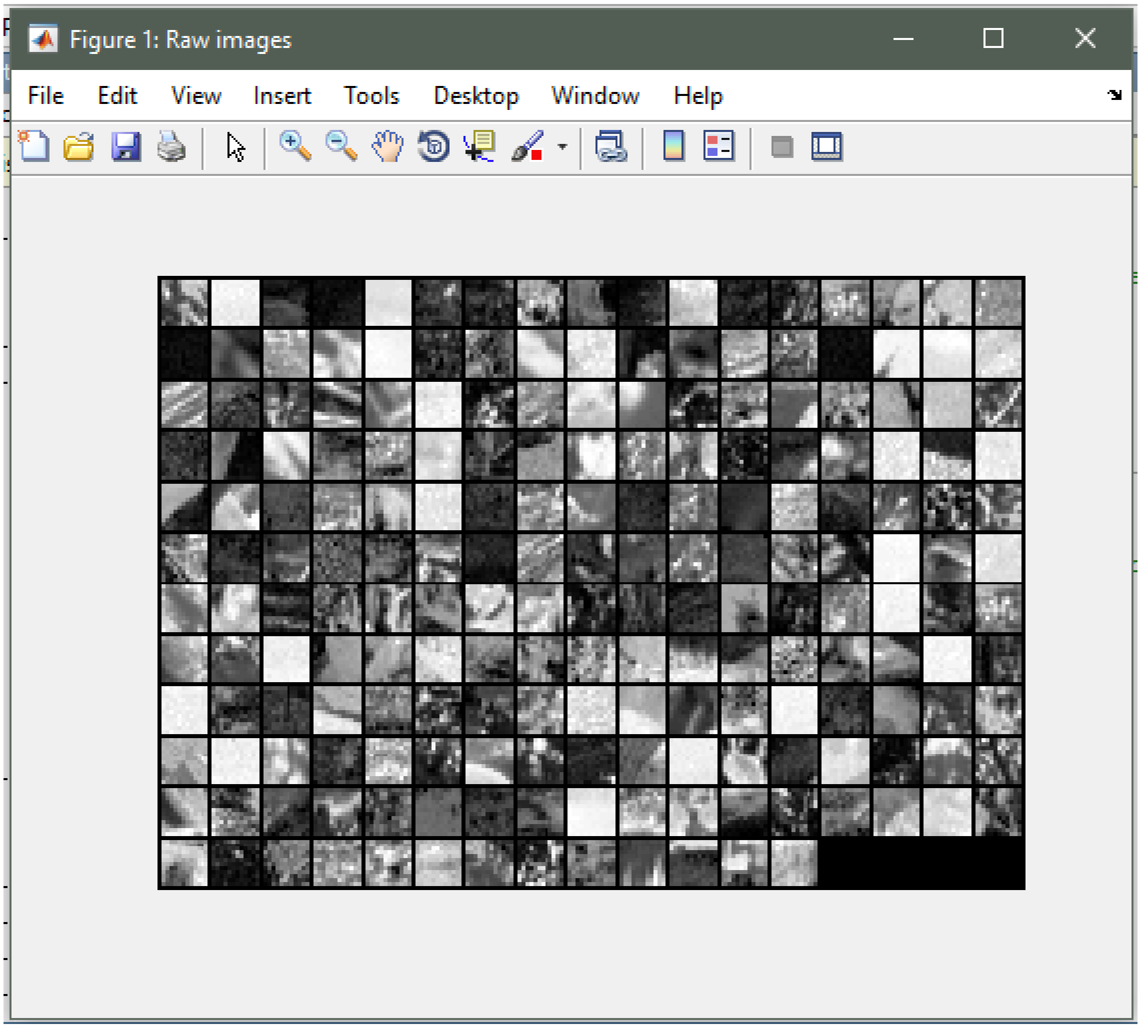
اگر بخواهیم x را از x سفید شده بدست آوریم:

$x_{Rot} \times U = x$ خود متوسط

$x_{white} \times U \longrightarrow$ تقریبی از x را دارد.

PAPCO

شکل های همه بخش ها به ازای واریانس 90%:



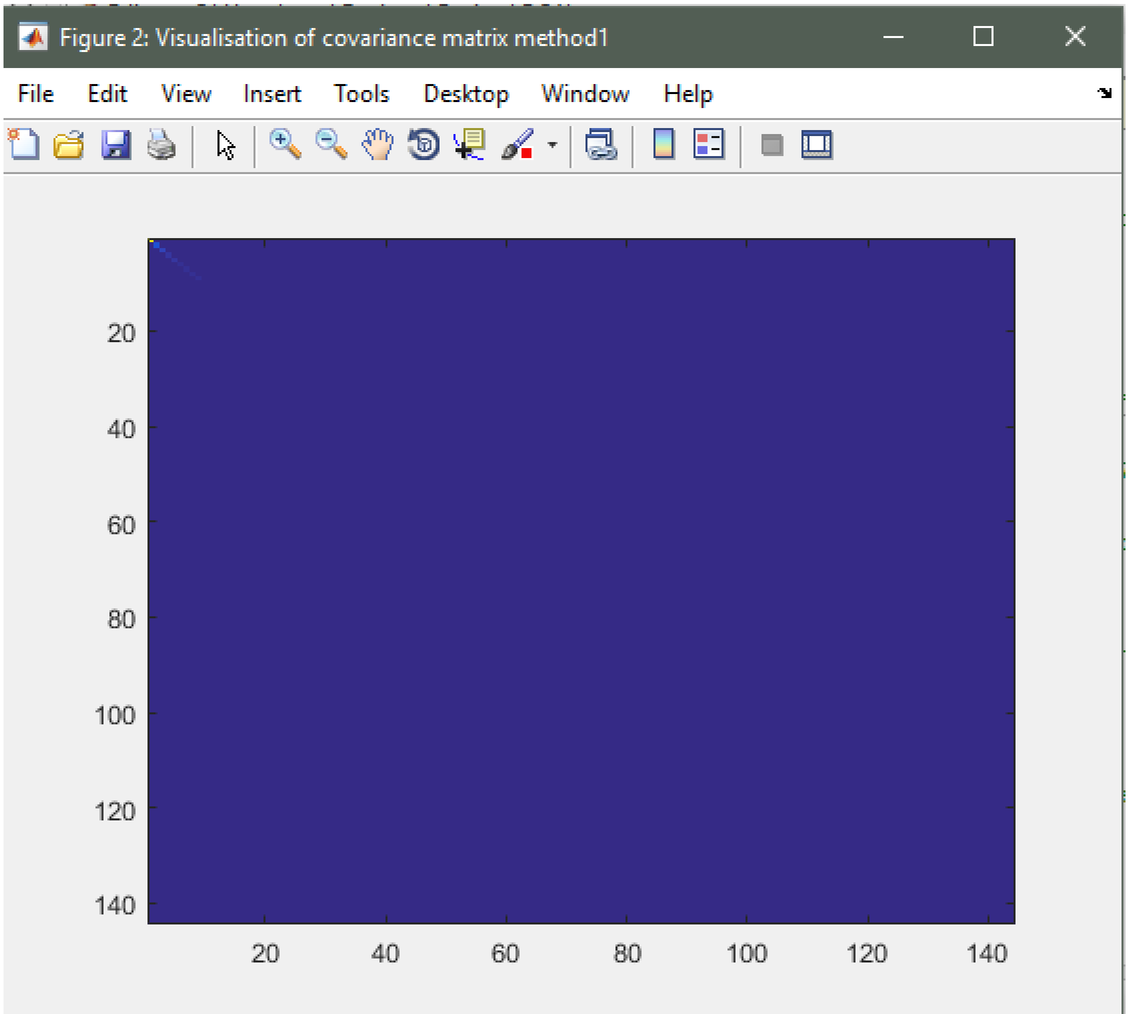


Figure 2: Visualisation of covariance matrix method1
File Edit View Insert Tools Desktop Window Help
20
40
60
80
100
120
140
20
40
60
80
100
120
140

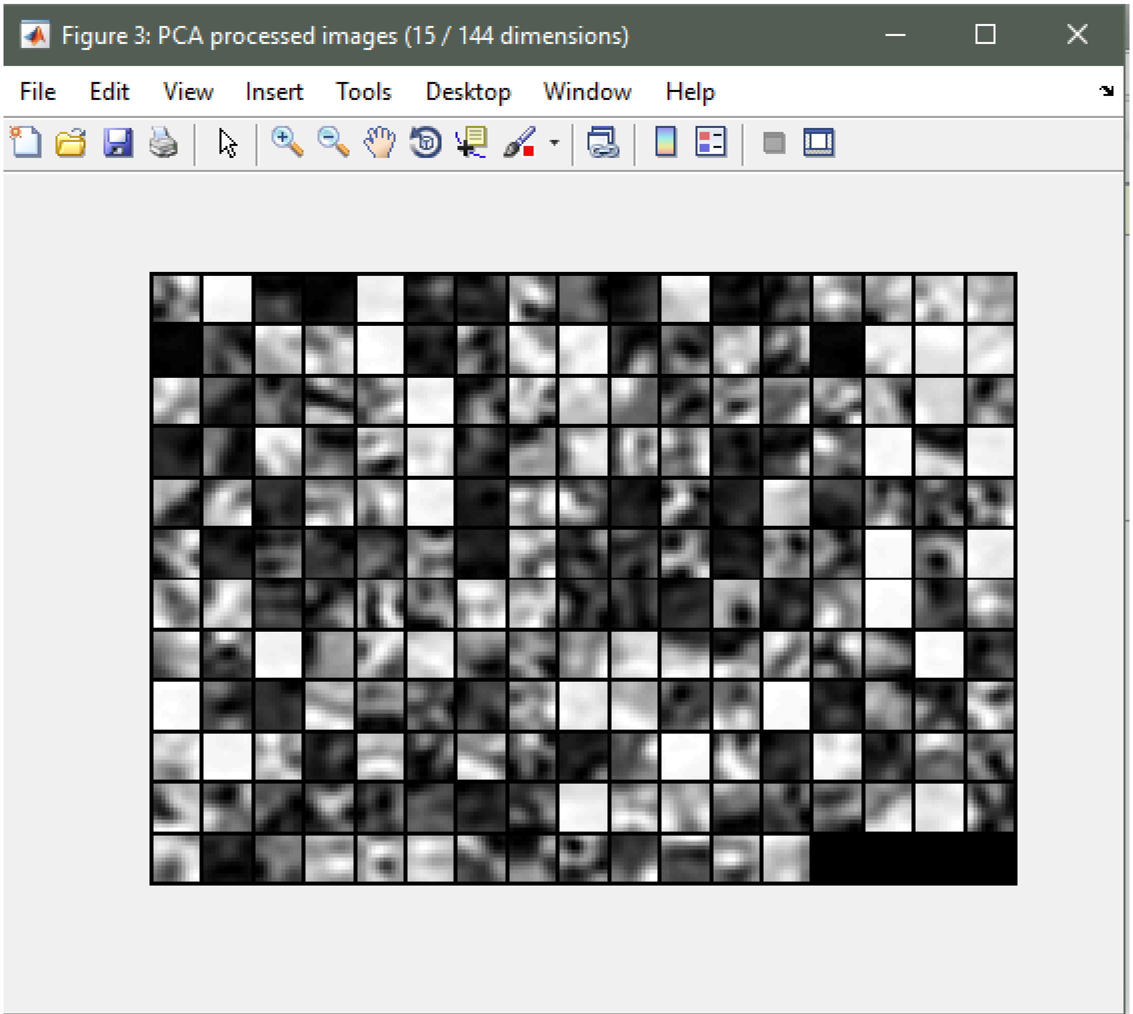


Figure 3: PCA processed images (15 / 144 dimensions)
File Edit View Insert Tools Desktop Window Help

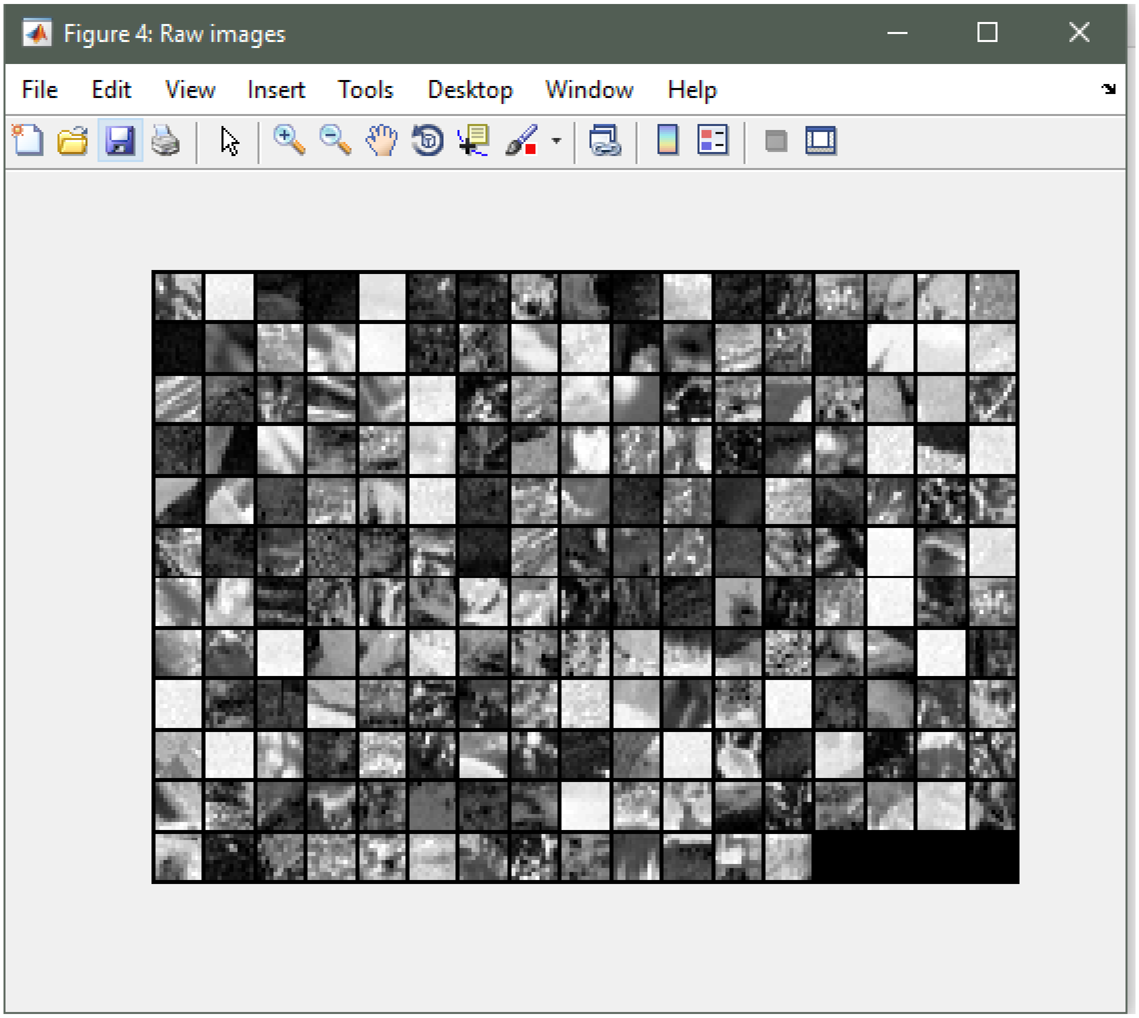


Figure 4: Raw images
File Edit View Insert Tools Desktop Window Help

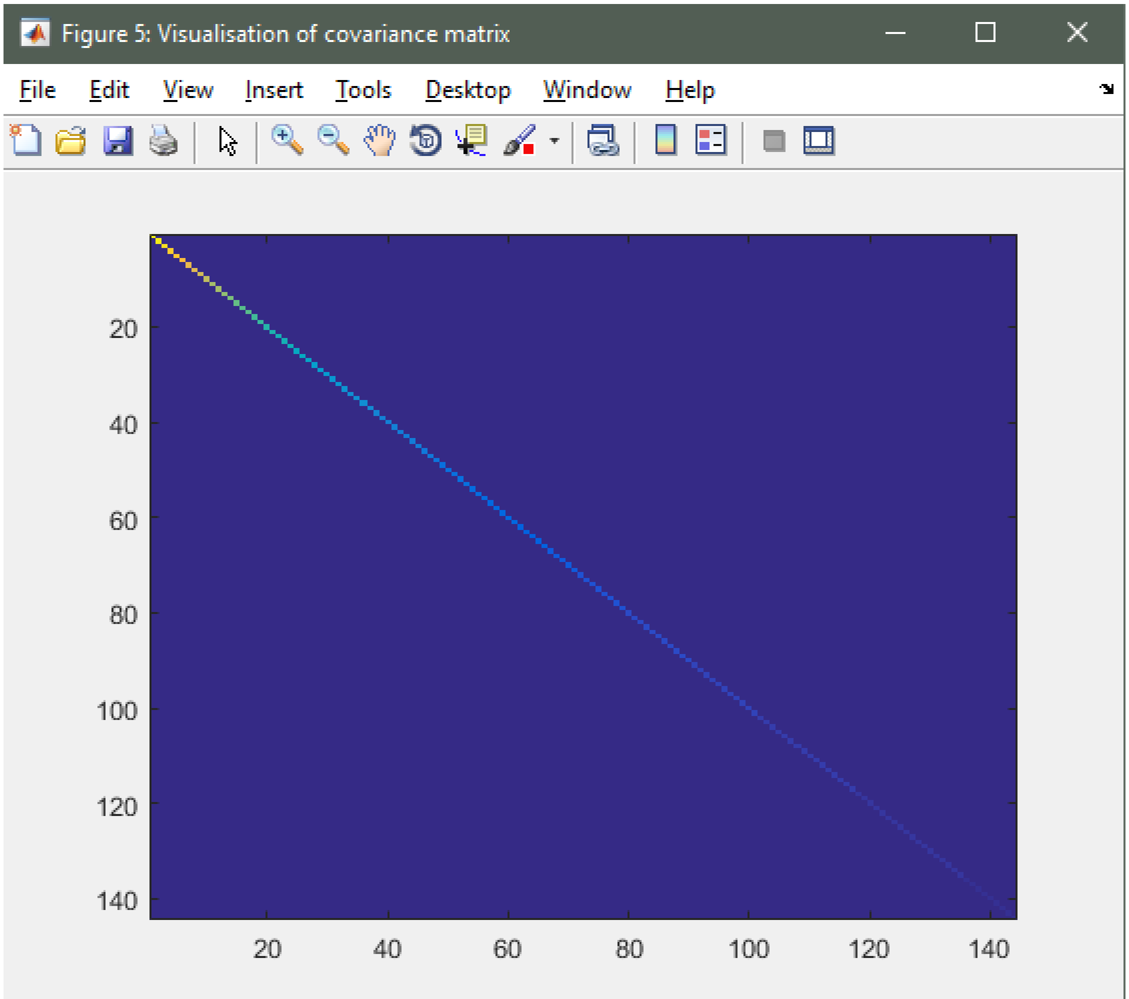


Figure 5: Visualisation of covariance matrix
File
Edit
View
Insert
Tools
Desktop
Window
Help
20
40
60
80
100
120
140
20
40
60
80
100
120
140

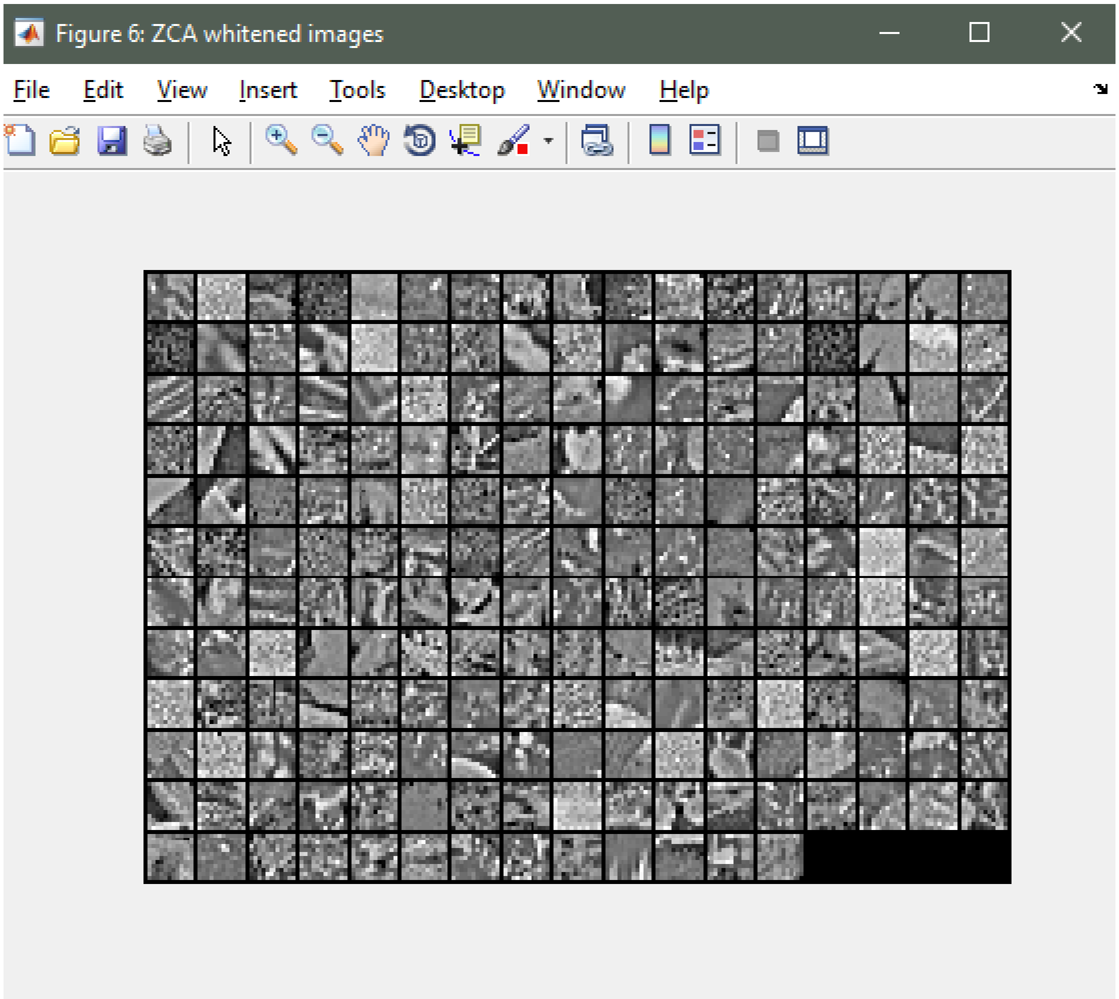


Figure 6: ZCA whitened images
File Edit View Insert Tools Desktop Window Help

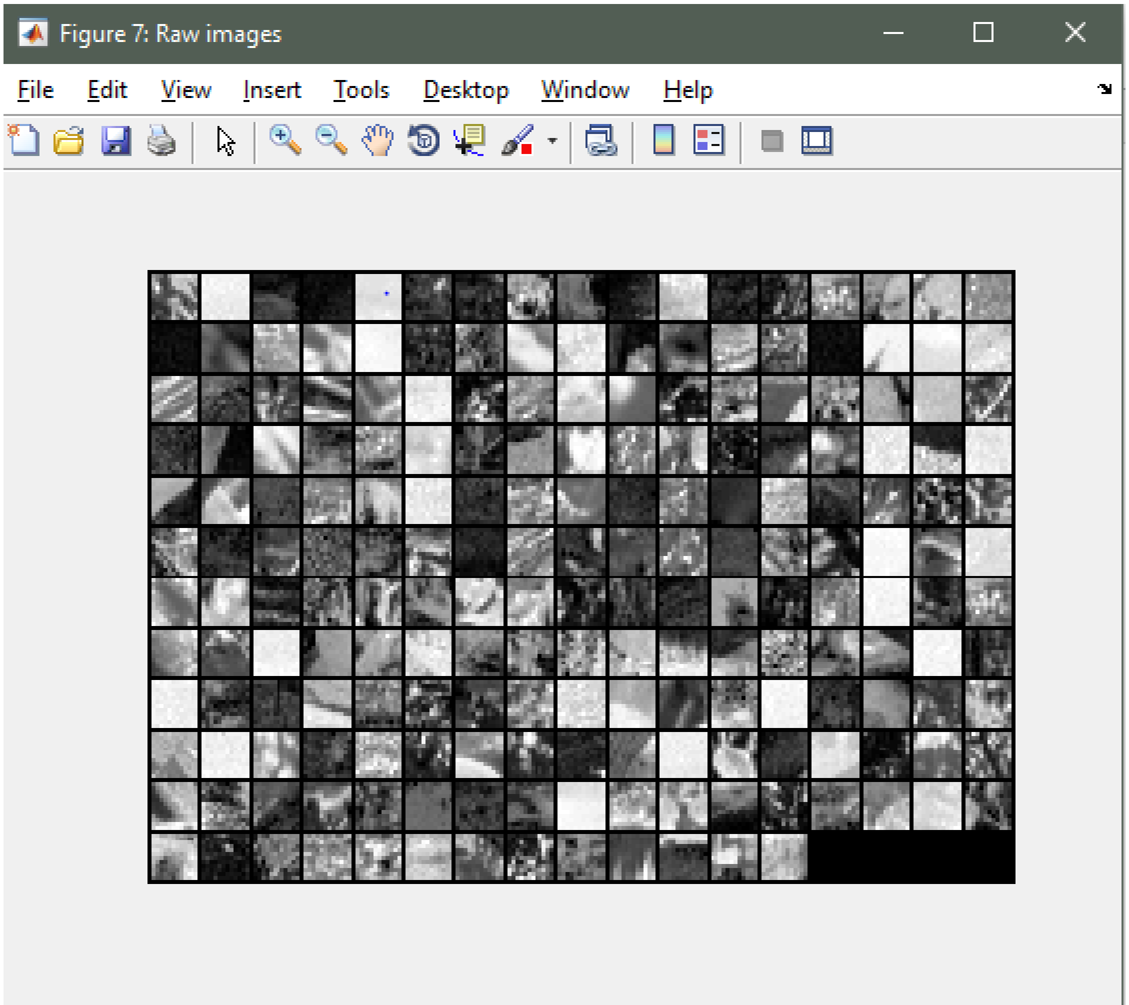
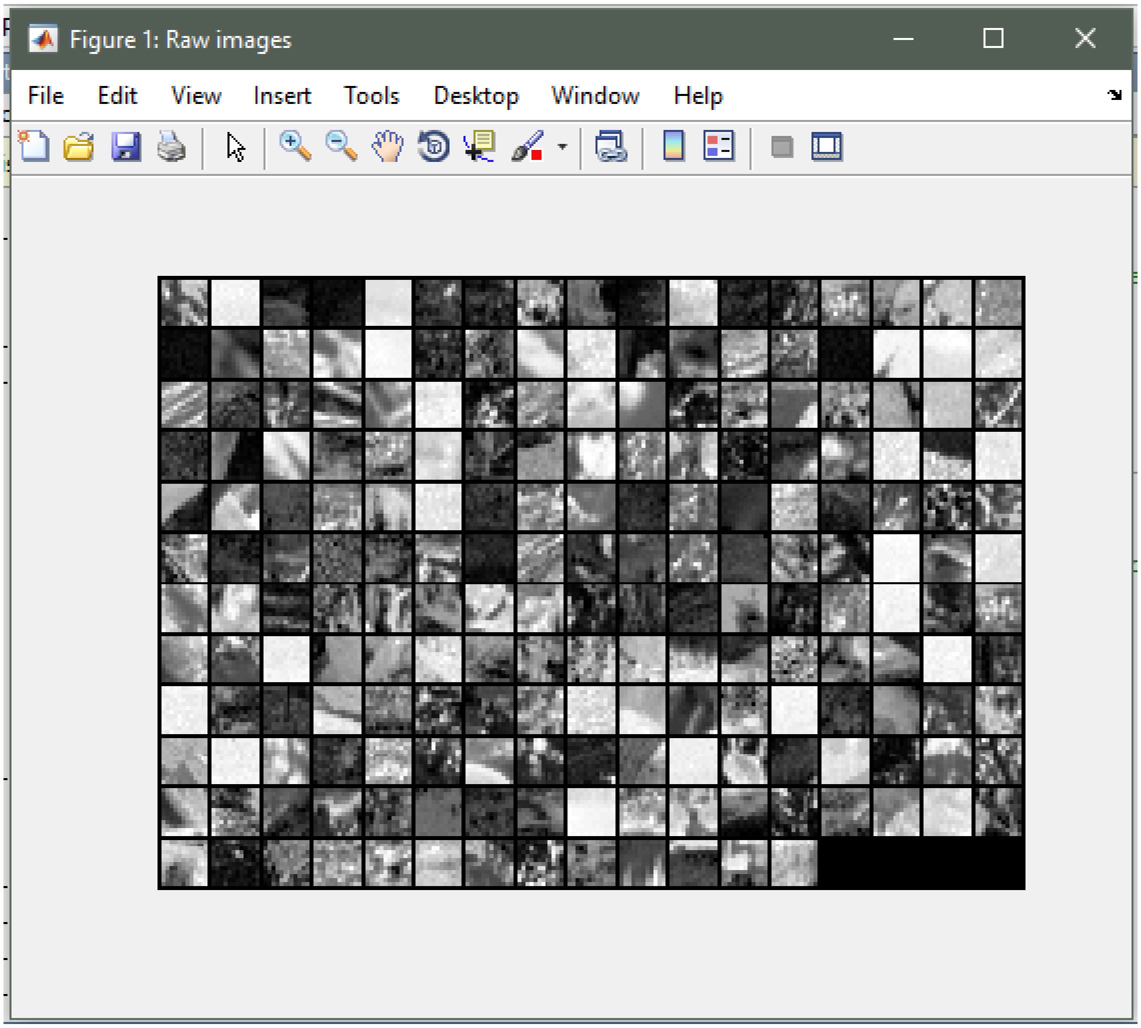


Figure 7: Raw images
File Edit View Insert Tools Desktop Window Help

شکل های تمام بخش ها به ازای واریانس ۹۹%



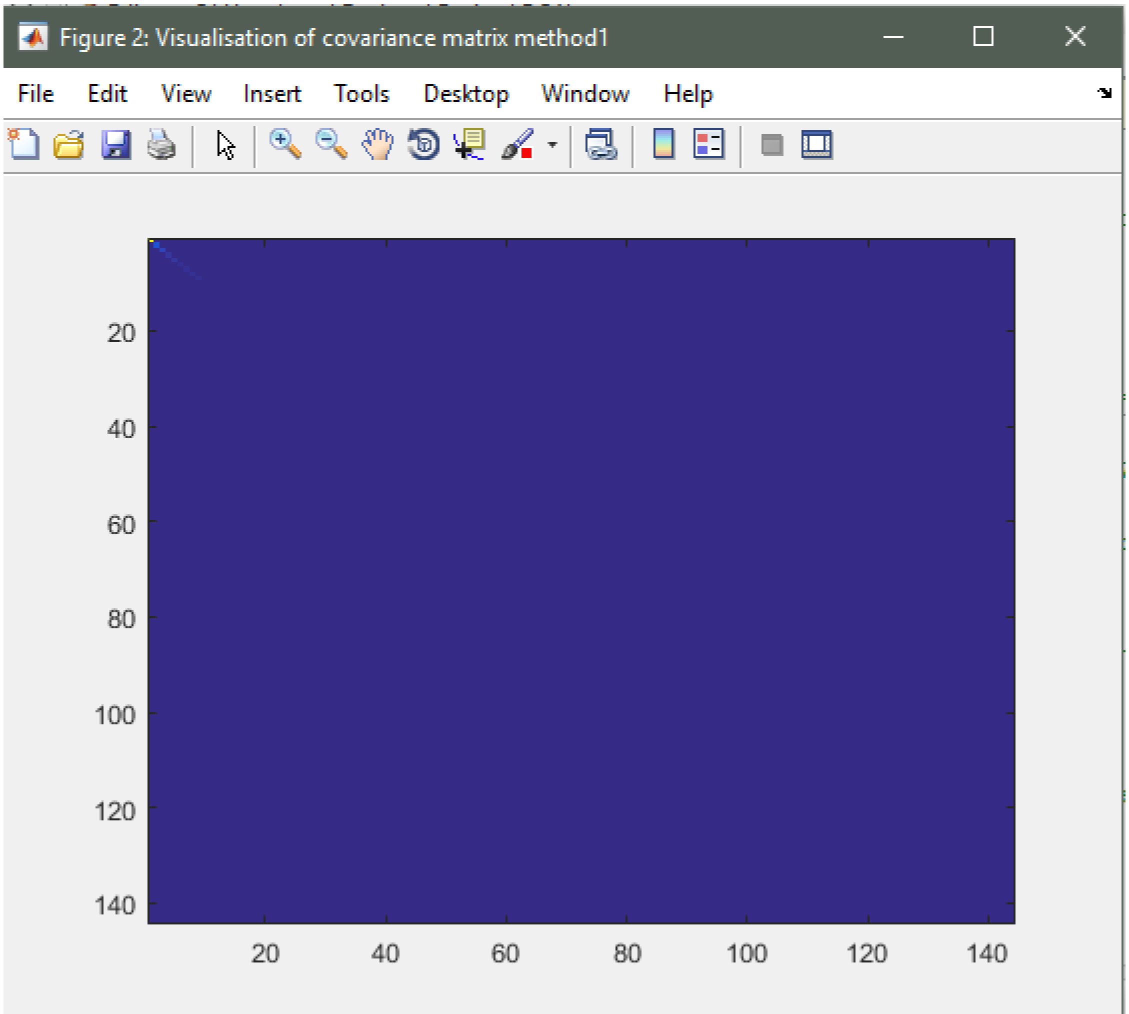


Figure 2: Visualisation of covariance matrix method1
File Edit View Insert Tools Desktop Window Help
20
40
60
80
100
120
140
20
40
60
80
100
120
140

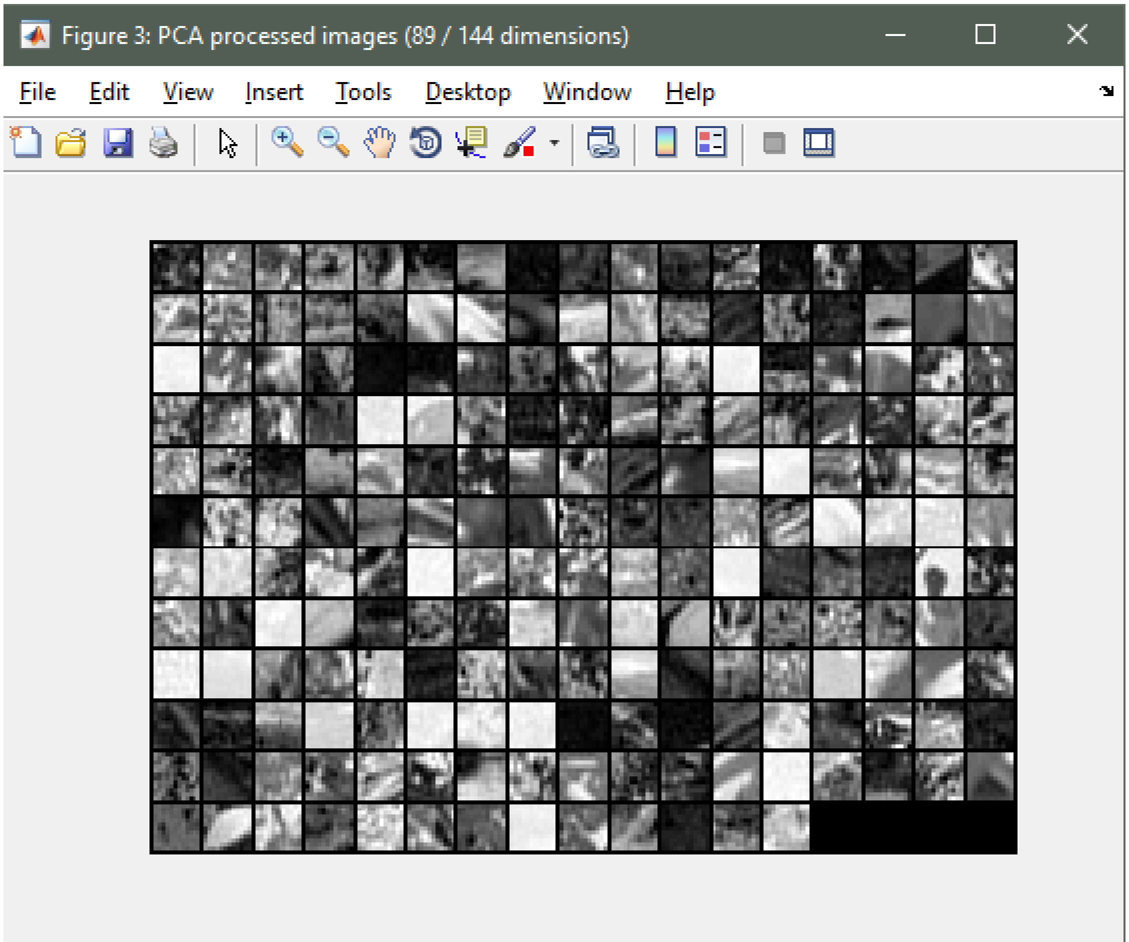


Figure 3: PCA processed images (89 / 144 dimensions)
File
Edit
View
Insert
Tools
Desktop
Window
Help

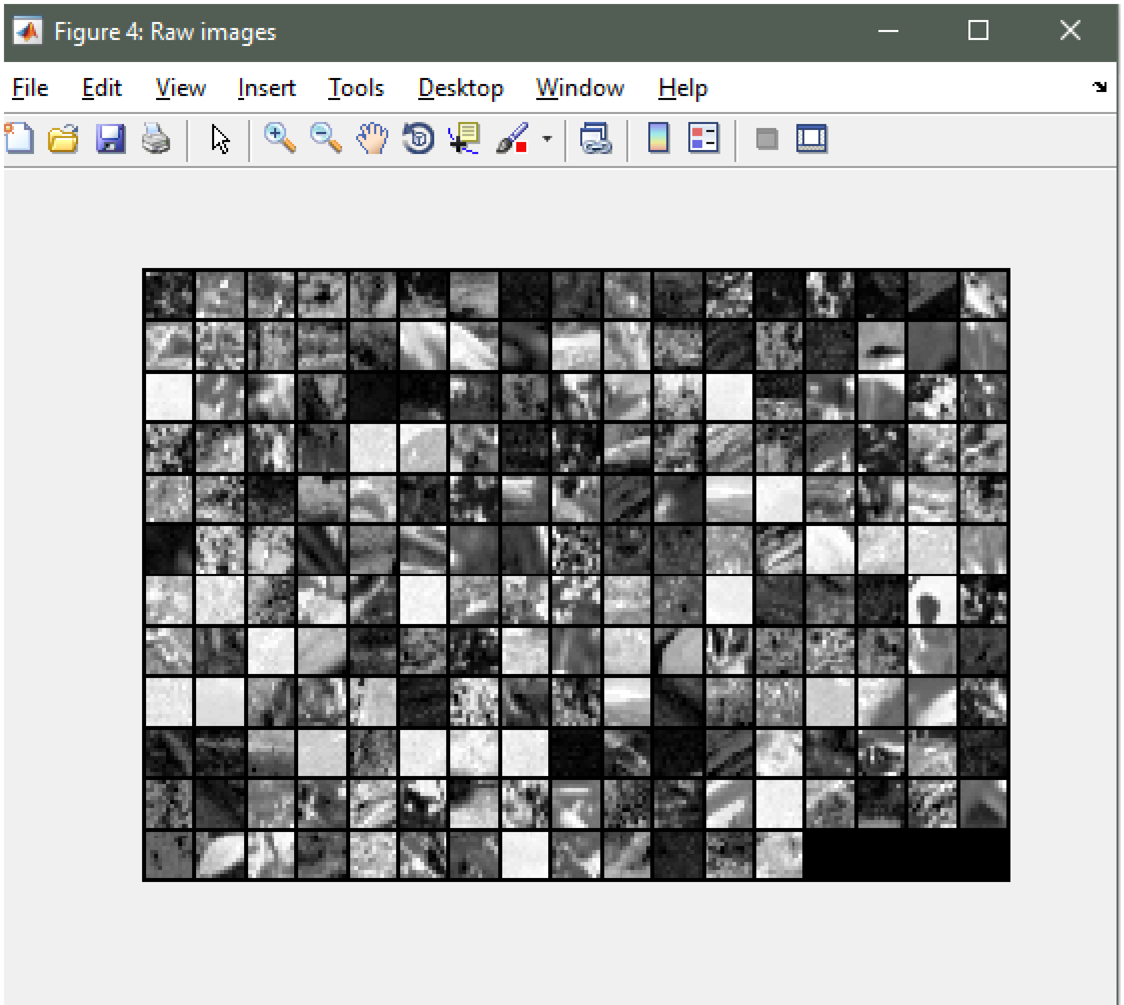


Figure 4: Raw images
File Edit View Insert Tools Desktop Window Help

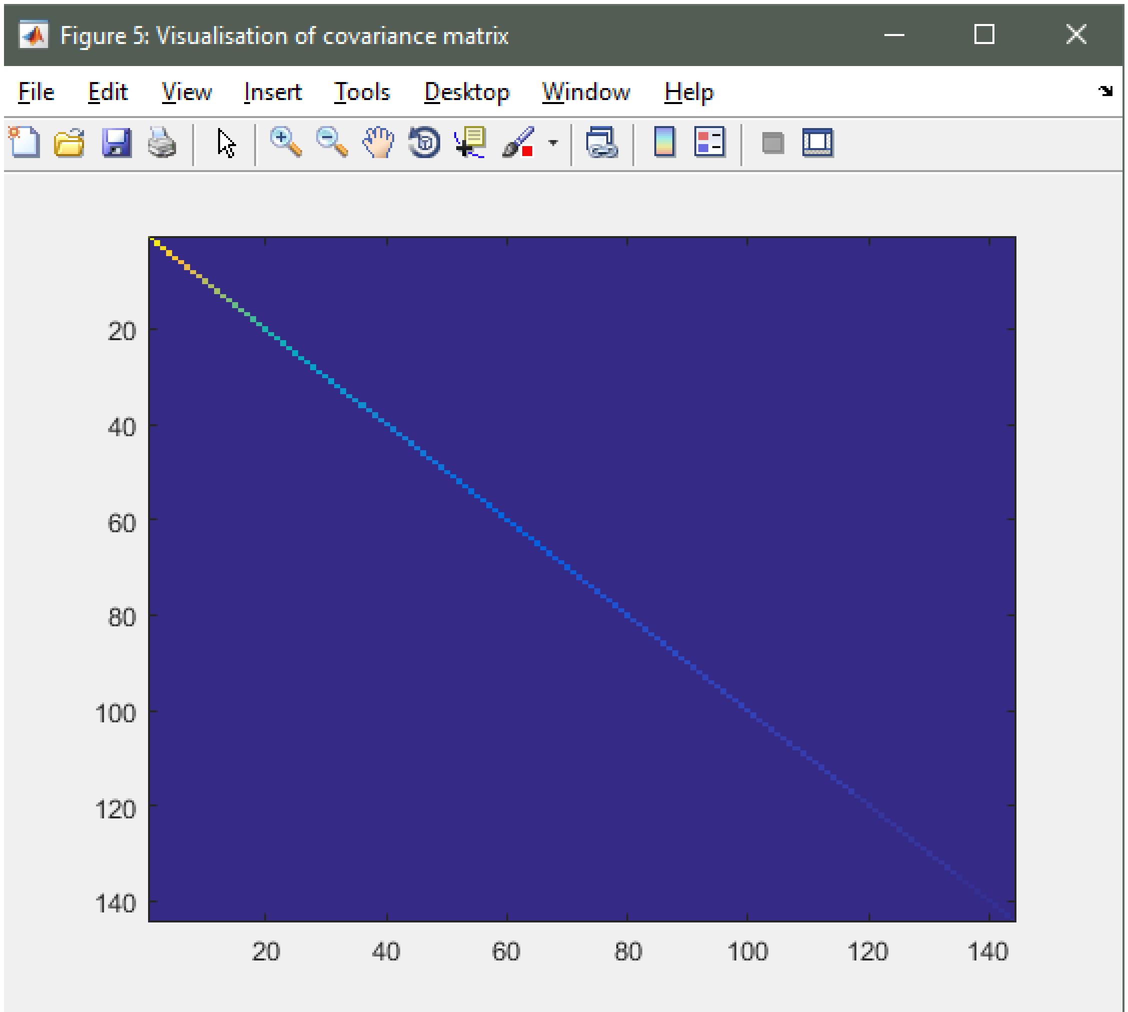


Figure 5: Visualisation of covariance matrix
File Edit View Insert Tools Desktop Window Help
20
40
60
80
100
120
140
20
40
60
80
100
120
140

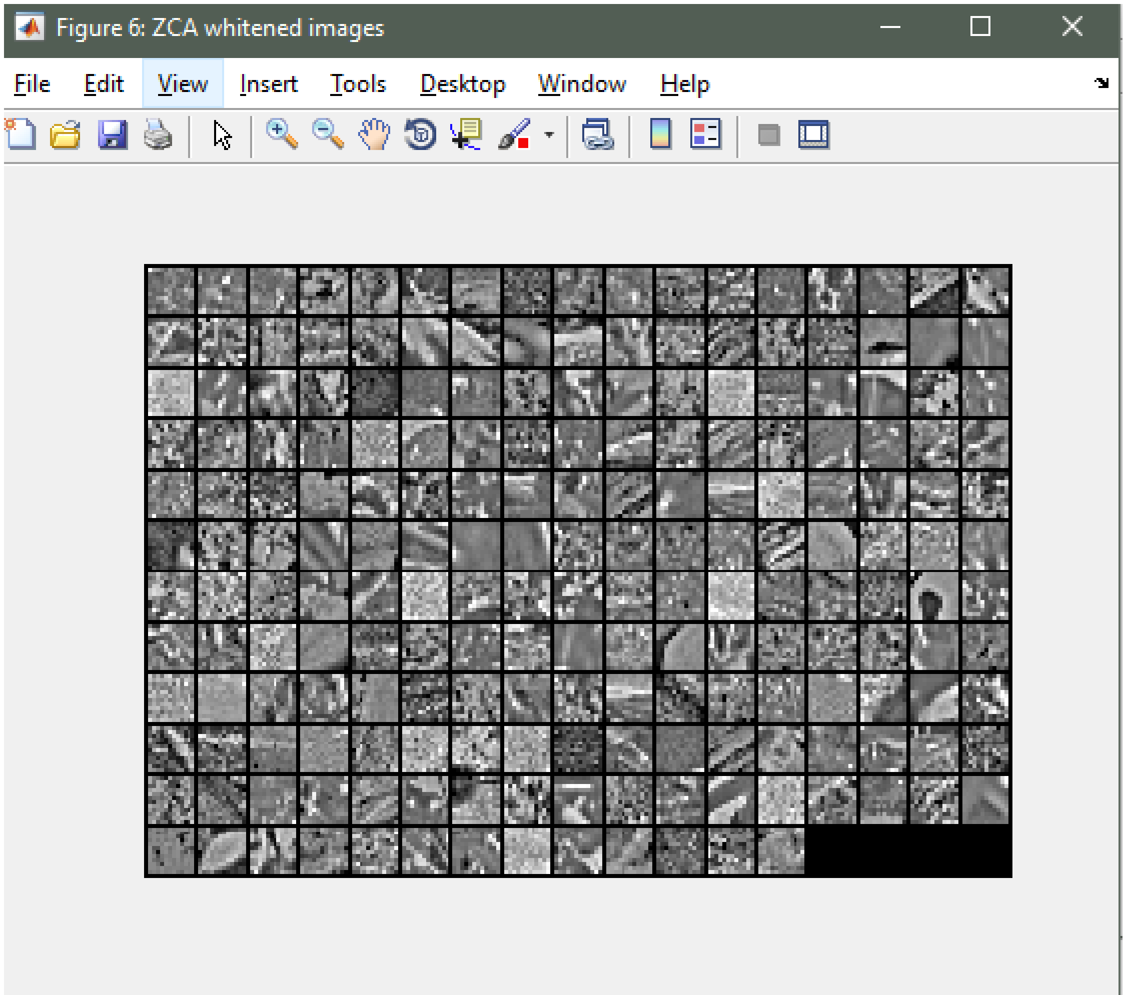


Figure 6: ZCA whitened images
File
Edit
View
Insert
Tools
Desktop
Window
Help

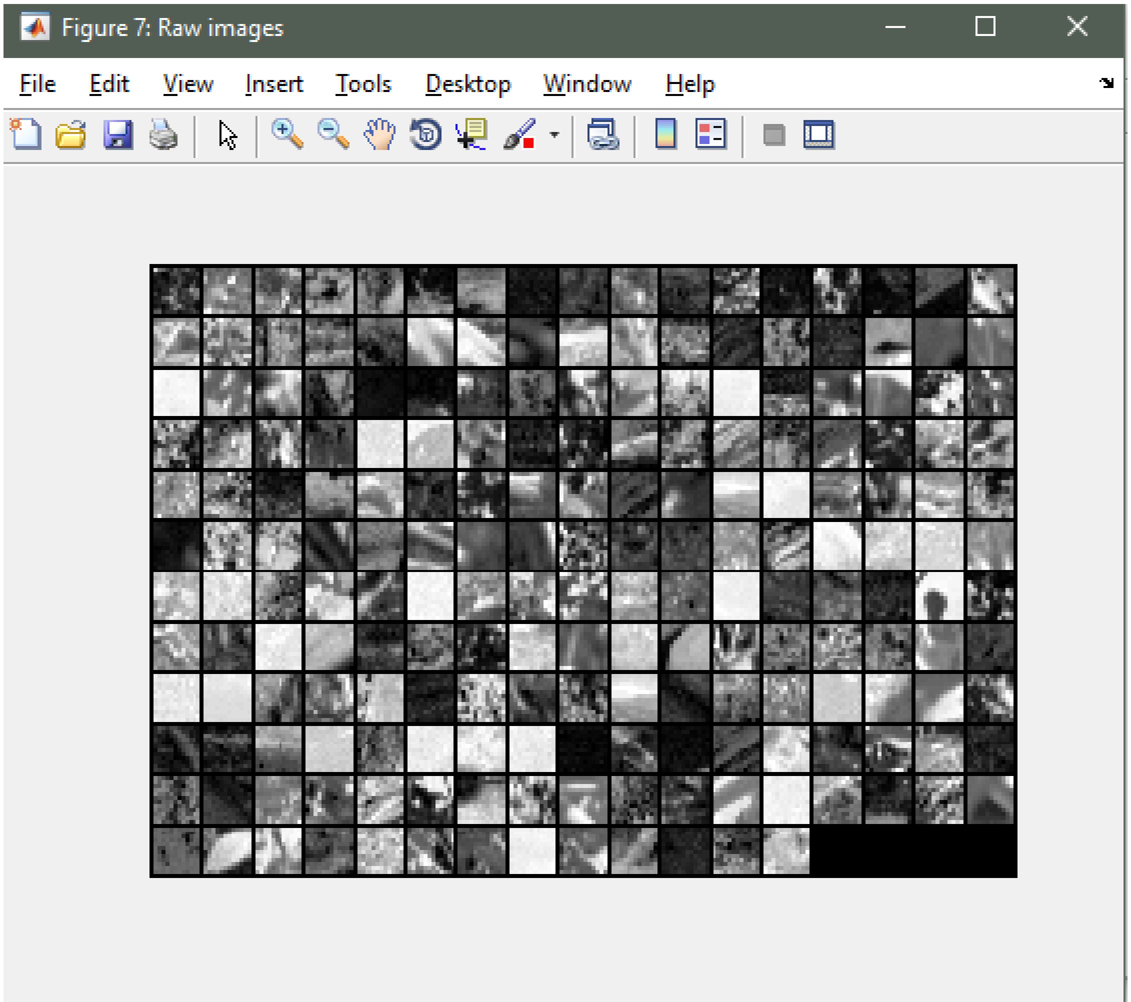


Figure 7: Raw images
File
Edit
View
Insert
Tools
Desktop
Window
Help