Project 2: SVD & PCA

Christopher Stoll & Michael Crouse

Singular Value Decomposition (SVD)

	M						=	U					*	S						*	Vt					
	1	2	3	4	5	6		1	2	3	4	5		1	2	3	4	5	6		1	2	3	4	5	6
1	6.2	6.2	6.2	6.2	6.2	6.2		1	1	1	1	1		3.2	0	0	0	0	0		1	1	1	1	1	1
2	6.2	6.2	6.2	6.2	6.2	6.2		1	1	1	1	1		0	1.6	0	0	0	0		1	1	1	1	1	1
3	6.2	6.2	6.2	6.2	6.2	6.2		1	1	1	1	1		0	0	0.8	0	0	0		1	1	1	1	1	1
4	6.2	6.2	6.2	6.2	6.2	6.2		1	1	1	1	1		0	0	0	0.4	0	0		1	1	1	1	1	1
5	6.2	6.2	6.2	6.2	6.2	6.2		1	1	1	1	1		0	0	0	0	0.2	0		1	1	1	1	1	1
6																					1	1	1	1	1	1
1								3.2	1.6	0.8	0.4	0.2	0													
2								3.2	1.6	0.8	0.4	0.2	0													
3								3.2	1.6	0.8	0.4	0.2	0													
4								3.2	1.6	0.8	0.4	0.2	0													
5								3.2	1.6	0.8	0.4	0.2	0													

Singular Value Decomposition (SVD)

- Implementation
  - C++ with Eigen library
  - Python with Numpy

SVD for Image Compression

	M						=	U					*	S						*	Vt					
	1	2	3	4	5	6		1	2	3	4	5		1	2	3	4	5	6		1	2	3	4	5	6
1	48.13	48.13	48.13	48.13	48.13	48.13		1	1	1	1	1		$\boxed{32}$	0	0	0	0	0		1	1	1	1	1	1
2	48.13	48.13	48.13	48.13	48.13	48.13		1	1	1	1	1		0	$\boxed{16}$	0	0	0	0		1	1	1	1	1	1
3	48.13	48.13	48.13	48.13	48.13	48.13		1	1	1	1	1		0	0	0.08	0	0	0		1	1	1	1	1	1
4	48.13	48.13	48.13	48.13	48.13	48.13		1	1	1	1	1		0	0	0	0.04	0	0		1	1	1	1	1	1
5	48.13	48.13	48.13	48.13	48.13	48.13		1	1	1	1	1		0	0	0	0	0.01	0		1	1	1	1	1	1
6								**10 values**						**2 values**							1	1	1	1	1	1
																					12 values					
1								32	16	0.08	0.04	0.01	0													
2								32	16	0.08	0.04	0.01	0													
3								32	16	0.08	0.04	0.01	0													
4								32	16	0.08	0.04	0.01	0													
5								32	16	0.08	0.04	0.01	0													

10 + 2 + 12 = 24 values stored

(instead of 5 * 6 = 30)

SVD for Image Compression

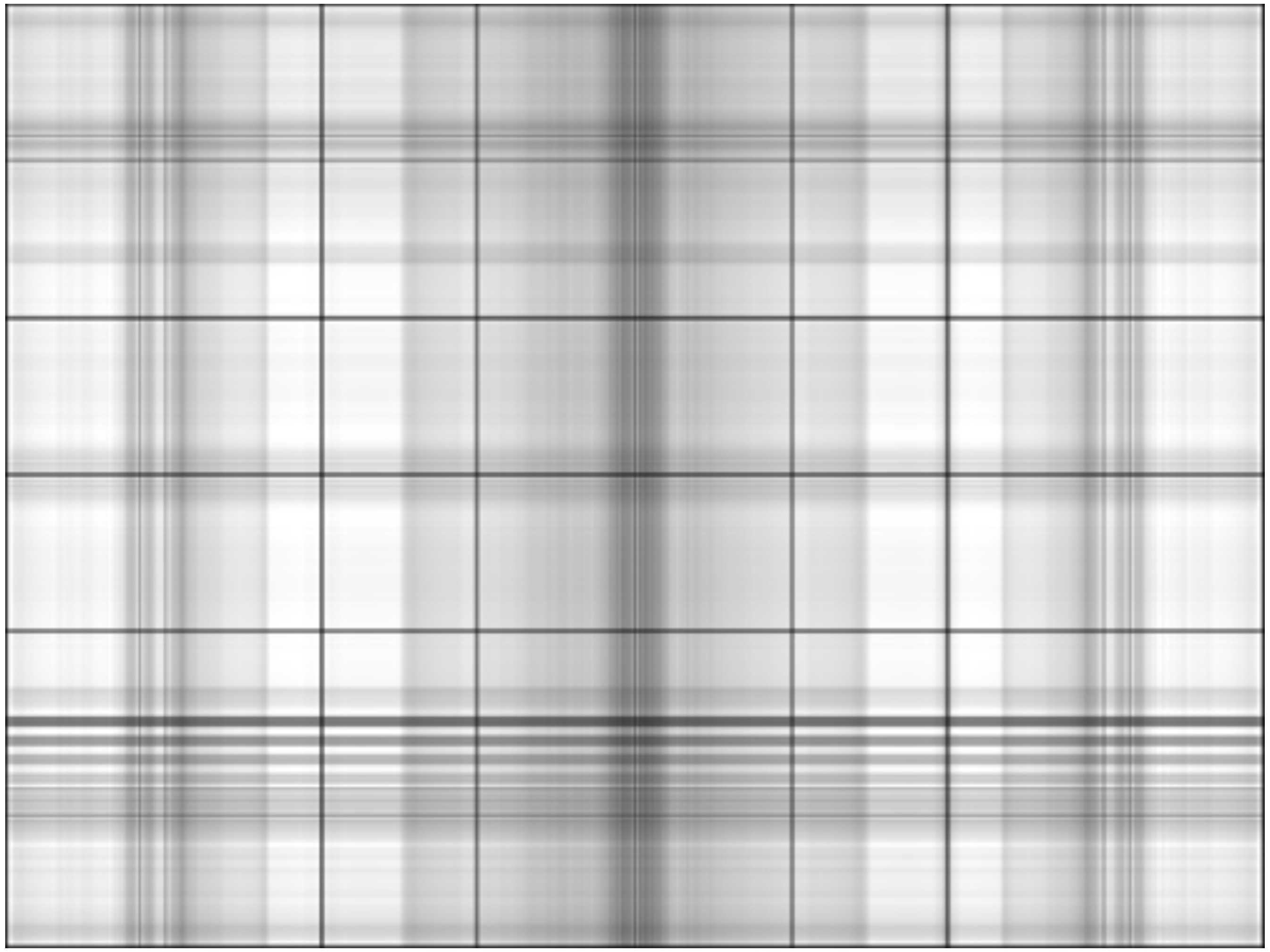
- Issues
  - PGM Pixel -- 1 byte
  - SVD Eigenvalues -- 8 byte double
  - SVD Eigenvector values -- 2 byte float

SVD for Image Compression

- "Compressed" image LARGER than uncompressed

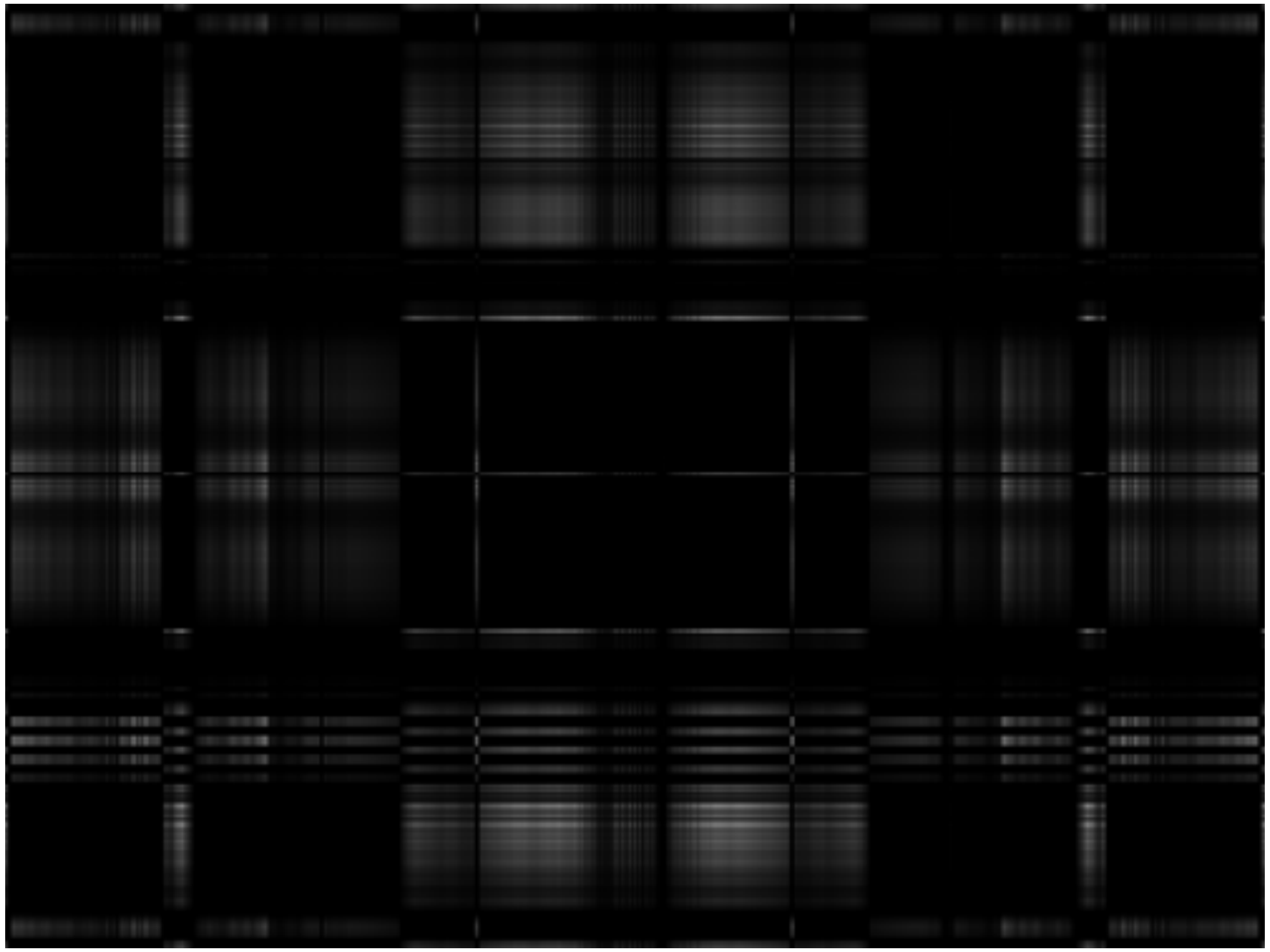
SVD for Image Compression Example 1

- The first eigenvalue



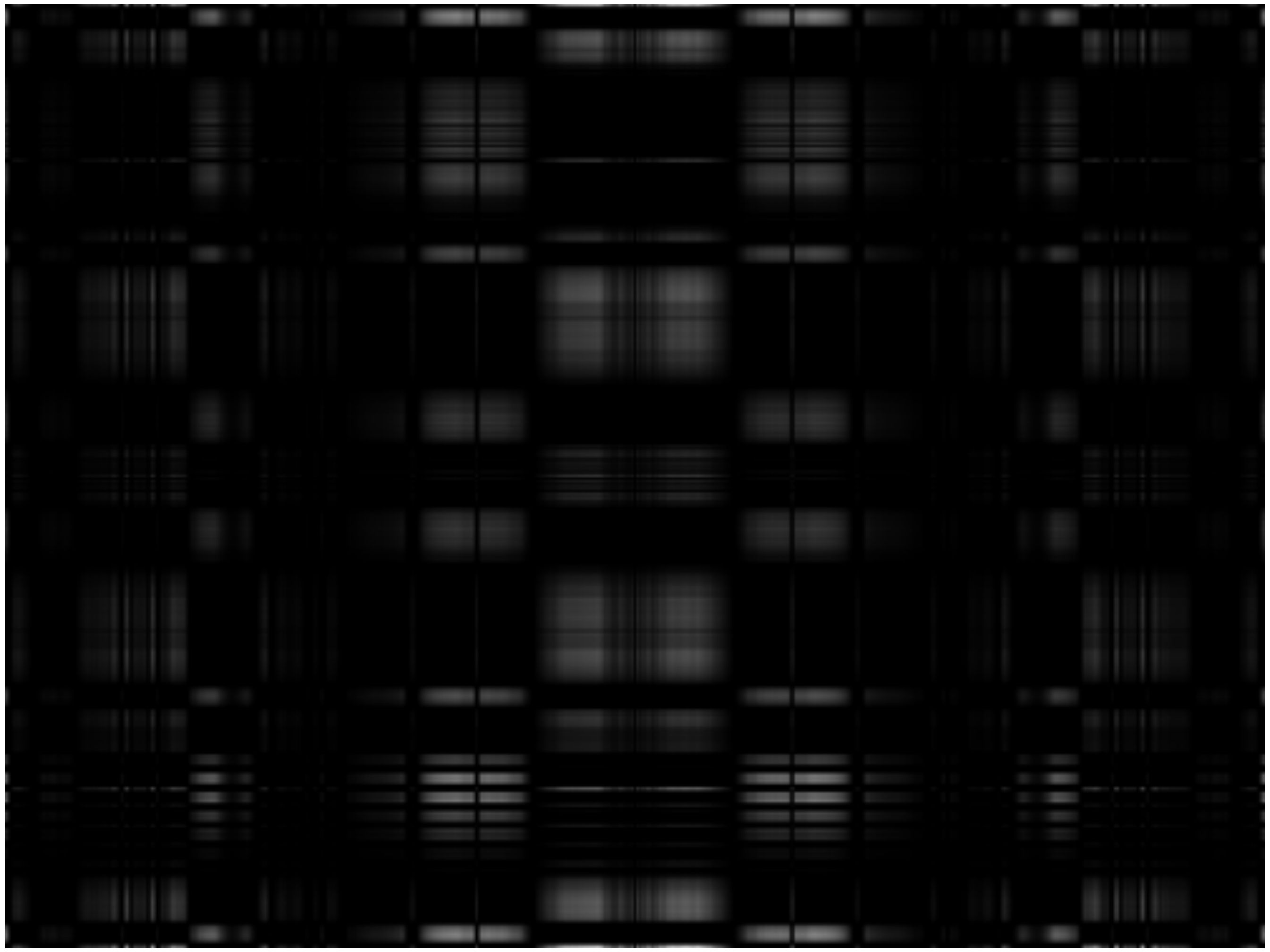
SVD for Image Compression Example 1

- The second eigenvalue



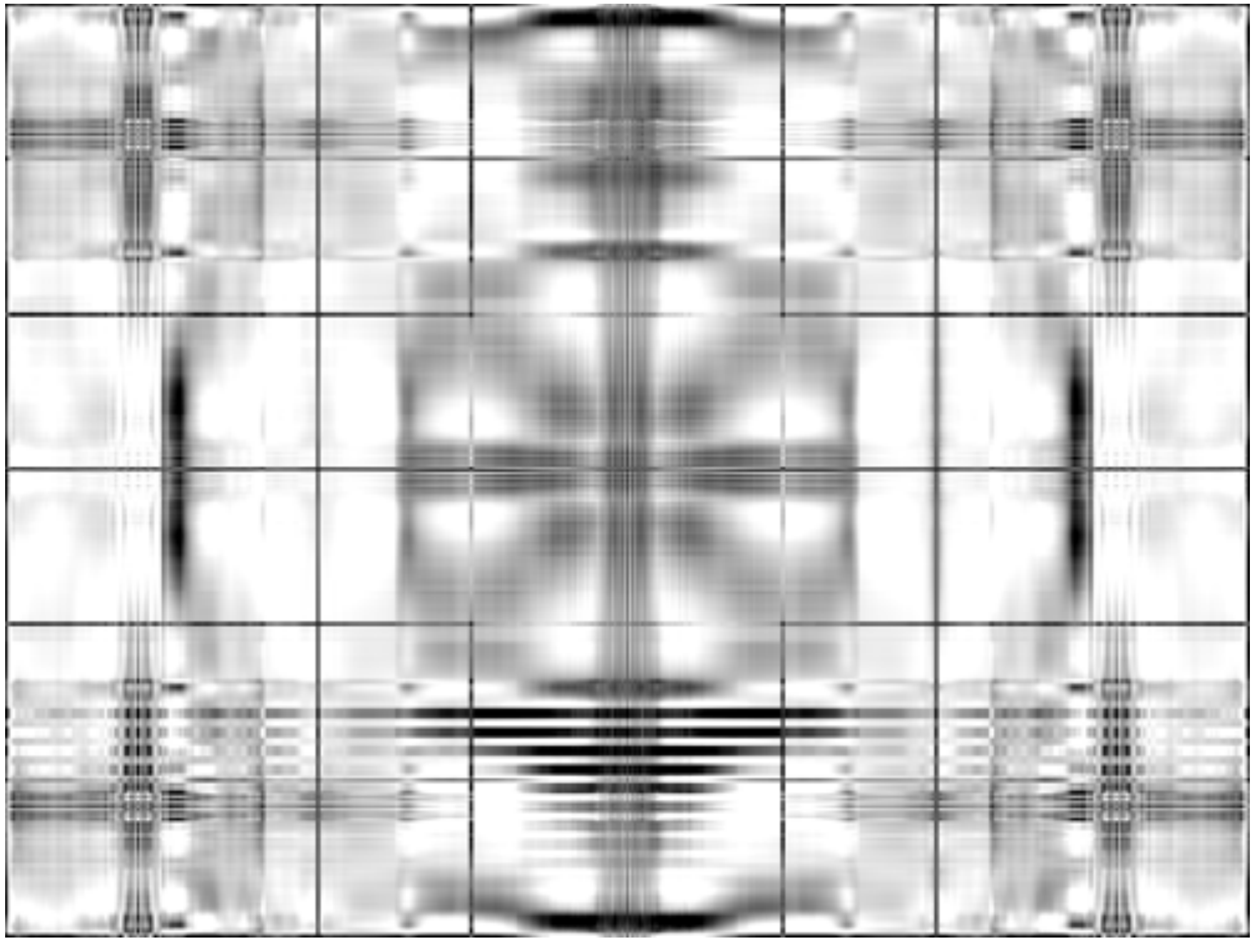
SVD for Image Compression Example 1

- The third eigenvalue



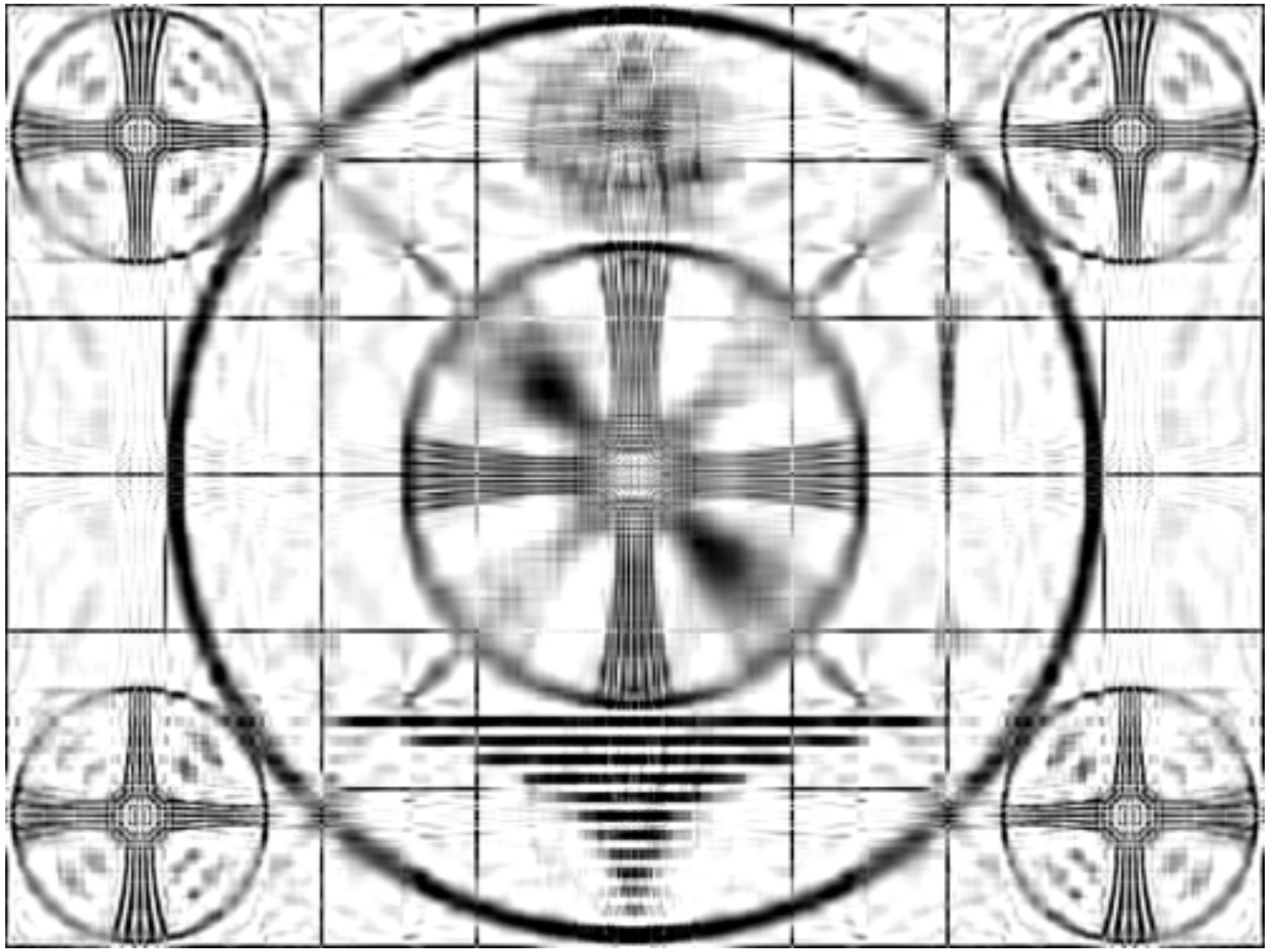
SVD for Image Compression Example 1

- The first 4 eigenvalues combined



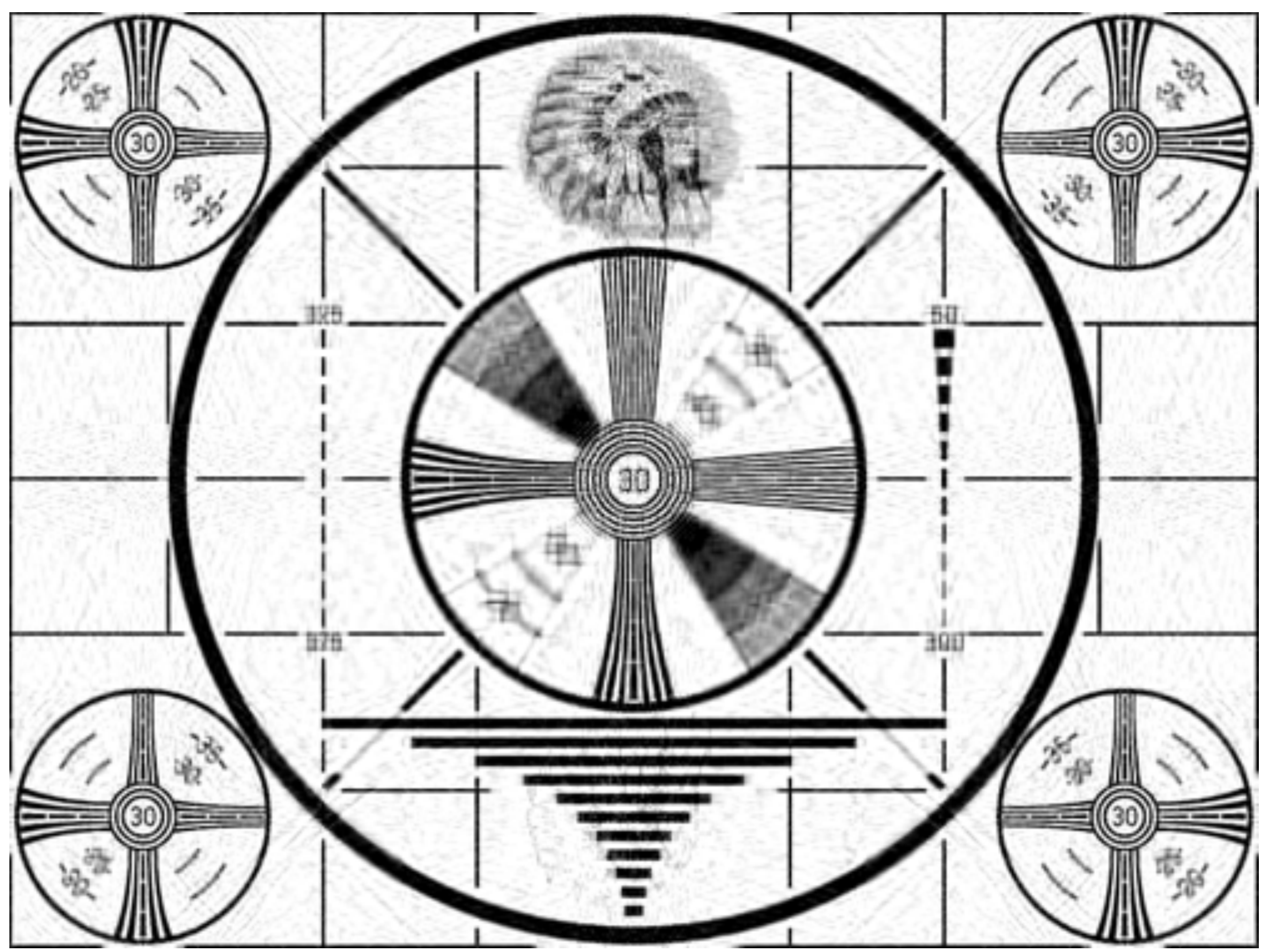
SVD for Image Compression Example 1

- The first 16 eigenvalues combined



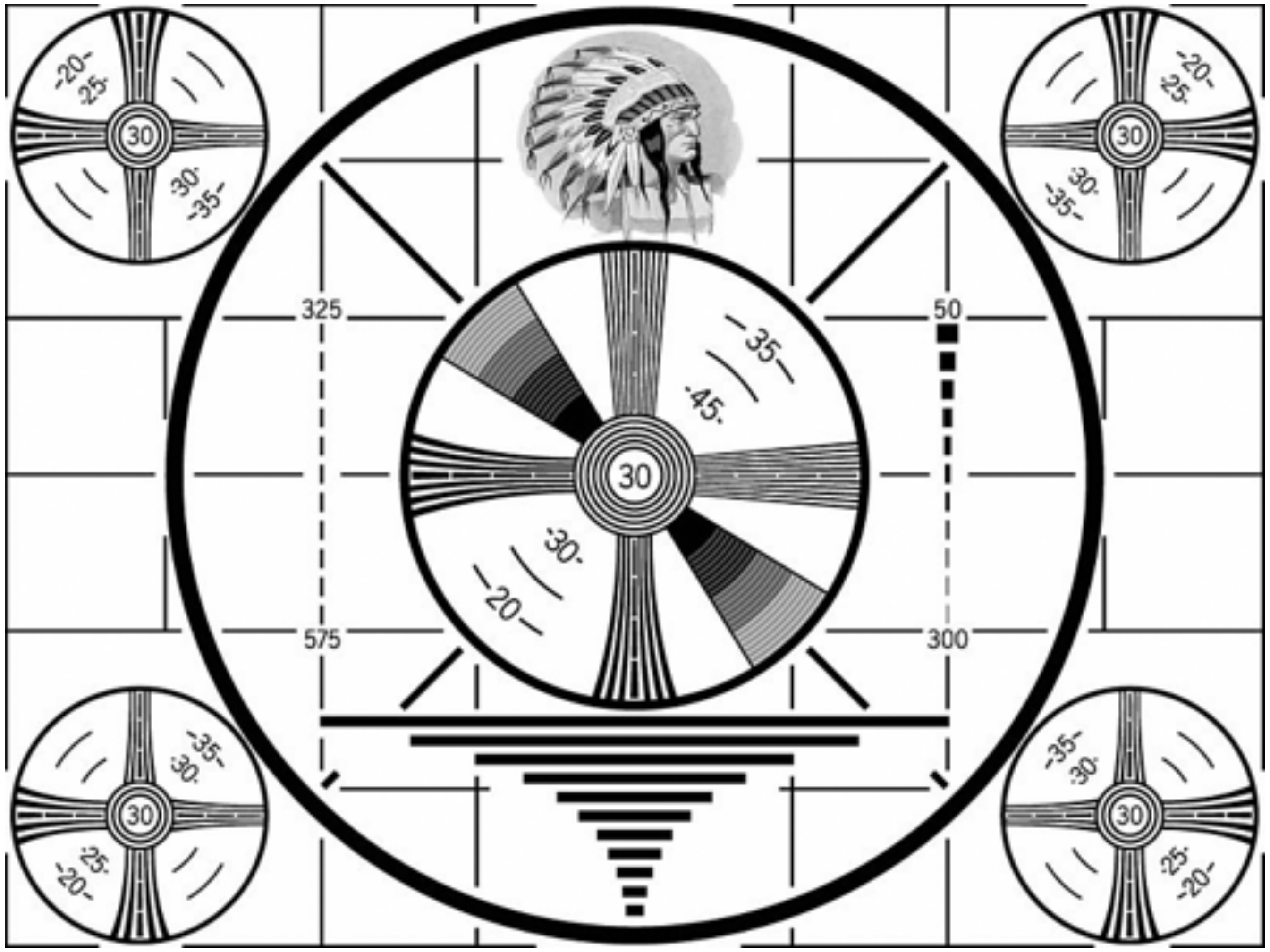
SVD for Image Compression Example 1

- The first 64 eigenvalues combined



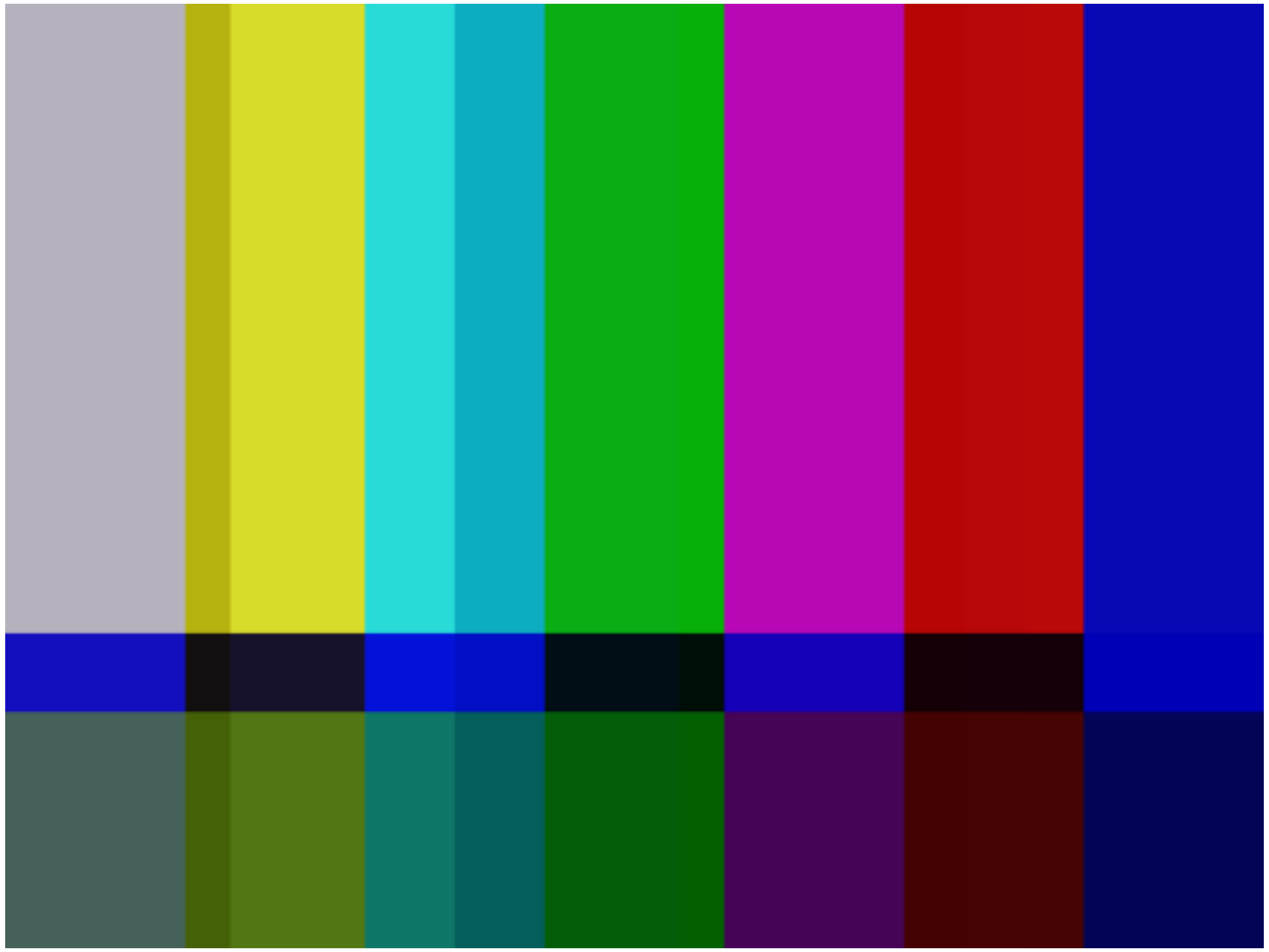
SVD for Image Compression Example 1

- The original image



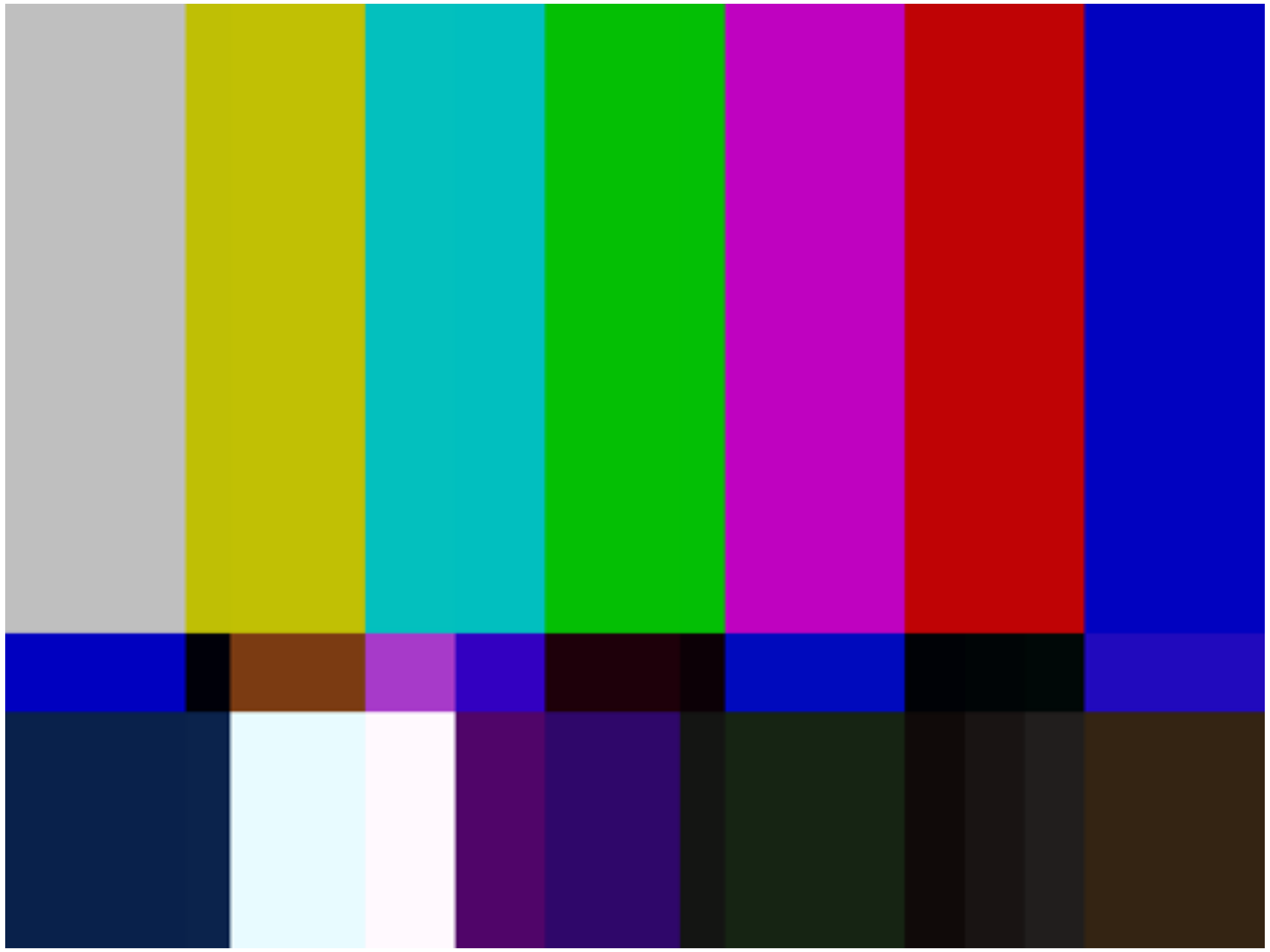
SVD for Image Compression Example 2

- The first eigenvalue



SVD for Image Compression Example 2

- The first 2 eigenvalues combined



SVD for Image Compression Example 2

- The first 3 eigenvalues combined

SVD for Image Compression Example 2

- The original image

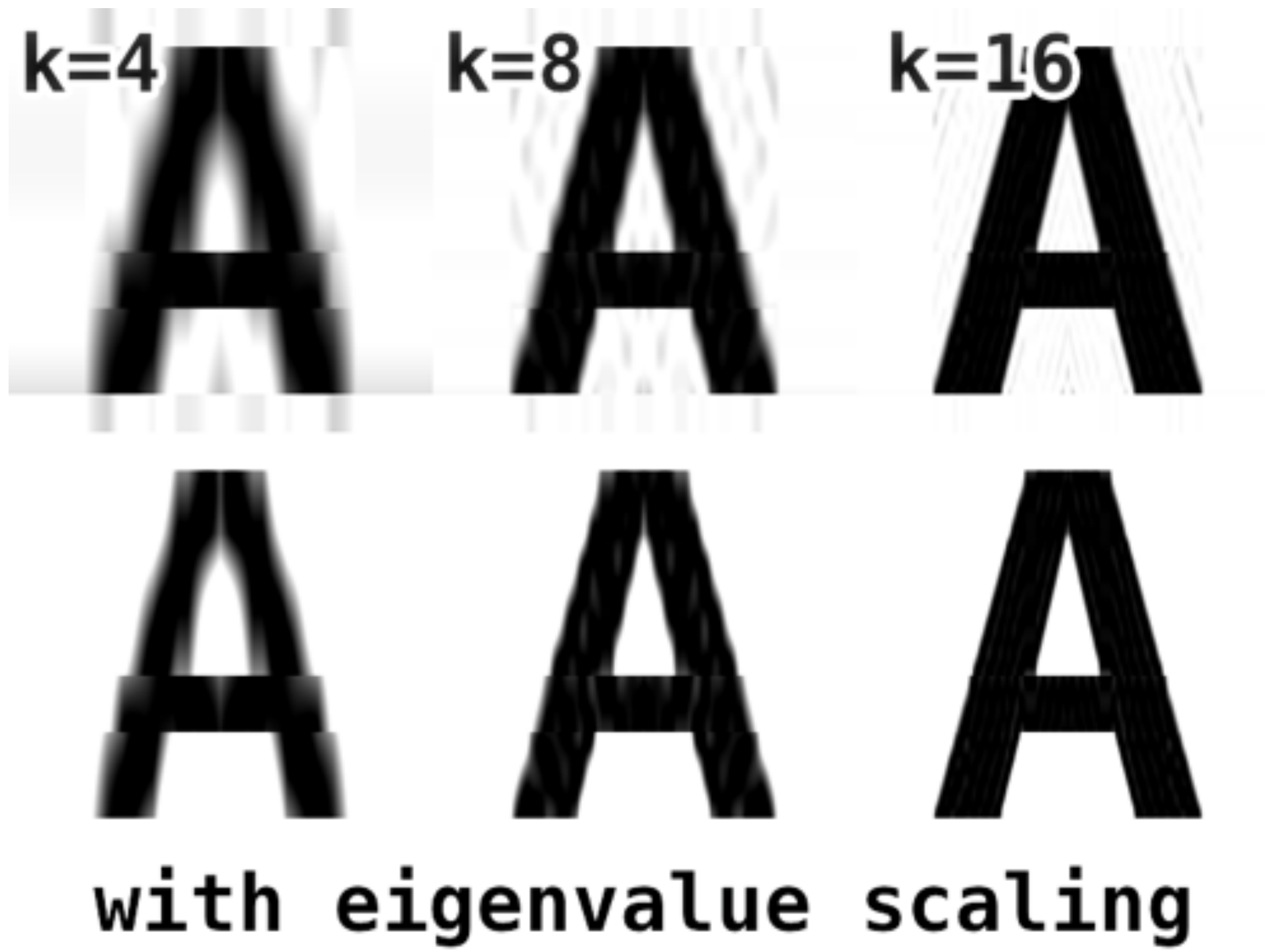
SVD for Image Compression

- Example 1
  - 800x600
  - k > 256 to be indistinguishable from original
- Example 2
  - 672x504
  - k >= 3 to be indistinguishable from original

SVD on High Contrast Images

- With B&W most of the white comes from the top eigenvalues
- Can take advantage of this to reduce artifacts
- Scale each eigenvector by a factor of k

SVD on High Contrast Images



SVD Empirical Results

	Size	Original	Binary	Rate	Cmprs Size	Rate
A	200x115	88K	22K	75%	19K	14%
B	400x300	423K	117K	72%	41K	65%
C	400x300	429K	117K	73%	41K	65%

Principal Component Analysis (PCA)

- Rotate mean-centered data around a new set of axes — the principal components
- Project data to lower dimensional space
- Removes redundant information
- Use SVD on data to determine the principal components

PCA Data Set One

- Barberton Police Records
  - March 2002 - February 2011
  - People involved in incidents
    - Demographic information
    - Count of involvement types

PCA with Weka

	V1	V2
dobYear	0.3871	0.144
height	0.3255	-0.0718
race=0.0	-0.4333	-0.2112
race=W	0.3543	0.3232
race=U	-0.0002	-0.0166
race=B	0.0139	-0.1981
race=A	-0.0069	-0.007
race=K	0.0004	-0.0244
race=M	-0.0064	-0.0028
race=L	-0.0001	-0.02
race=I	-0.0004	-0.0066
race=P	-0.003	-0.0095
race=N	-0.002	-0.0056

	V1	V1
gender=0.0	-0.4351	-0.2022
gender=M	0.2065	-0.2858
gender=F	0.045	0.411
gender=U	0.001	-0.0101
medicalCount	0.0608	0.015
offenseCount	0.1162	0.0374
warrantCount	0.2222	-0.4074
lastWarrantBail	0.1677	-0.324
involveCountVictim	0.0433	-0.0072
involveCountSuspect	0.194	-0.2922
involveCountArrestee	0.2391	-0.3668
involveCountComplain	0.0725	0.0149

PCA with Weka

- First principal component race=B is 0.0139 / race=W is 0.3543
  - The ratio is about .039
  - The B/W population ratio is about .065
- First principal component gender=M is 0.2065 / gender=F is 0.045
  - Notes
    - dobYear and height over-represented by 0 values (unknown)
    - race=0.0 means race is not recorded, often the victim

PCA with Weka