Waterfall Decomposition: Application of Truncated Singular Value Decomposition to Isolating Video Elements

Peter VandeHaar and Leah Wang

University of Michigan, Ann Arbor, MI

Abstract. Singular Value Decomposition (SVD) is widely regarded as one of the most important matrix factorization methods since its discovery in 1873 [5]. In the context of image analysis, it has been demonstrated that SVD can isolate the key elements within an image by using the raw pixel data. Additionally, images can be reconstructed using the largest few singular values from the SVD [1]. In our study, we extend the application of SVD to the separation of moving and stationary elements of a video. Since a video can be decomposed into a series of images with small changes between frames, we regard each pixel as a node within a network and track its changes throughout the frames of the video. We demonstrate that the largest singular values within a network are able to capture and separate the waterfall, splashes, and background stationary elements within the video. This investigation highlights the ability of SVD in separating elements with varying movements within a video, suggesting its potential utility in video analysis and object recognition.

1 Introduction

Singular Value Decomposition (SVD) is regarded as one of the most powerful and commonly used tools in modern matrix computation with a wide range of interesting applications [3]. One of the key functionalities of SVD is the reduction of large matrices into a smaller invertible and square matrix [2]. Additionally, if the singular values are ordered in decreasing order and truncated after r terms, the rank-r approximation can reconstruct an image that is close to the original, with larger r corresponding to higher similarity [3] [2].

Our investigation focuses on the application of SVD to videos with dynamic texture (DT) properties. These videos are spatially repetitive and temporally varying with visual patterns exhibiting certain stationary properties [6]. Some of the classic examples include flames, waterfall, escalators, fire and smoke, flowing water etc. We focus our investigation on the example of a waterfall, where the background, such as rocks, logs, and other environmental elements are stable and undergo minuscule changes across the frames, while the flowing water and the splashing drops are the most variable elements in the system. We hypothesize that by analyzing correlation and changes between pixels across different frames of the video, we can re-imagine the water flow as a network of pixels changing across time. We first aggregate the pixel information by unfolding the matrix in each frame and fitting them into one large matrix. We hypothesize that by performing SVD on this large matrix, we will be able to isolate the moving elements since they contribute to the largest amount of variation in the pixels.

2 Methods

2.1 Data Source and Pre-processing

The video we are using is of a waterfall, shot on an iPhone at 30 frames per second with a resolution of 1920×1080, which is 2 million pixels. It is available **??**here. In order to cut down the computational burden, we first cropped the video using Procreate so that the majority of the content remaining is the waterfall [4]. The resulting cropped video has a dimension of 942px by 530px, with 380 frames and DPI of 264. We further decreased the size of the final pixel by averaging every 2px × 2px square. Our final data consists of $f = 380$ frames, each frame containing $n = 265$ pixels per row and $p = 471$ pixels per column.

Since we want to perform SVD on a single large data to account for the changes of pixels over time, we unfolded the matrix by the rows such that each row has $n \times p = 124815$ entries, and the matrix has $f = 380$ rows corresponding to each frame.

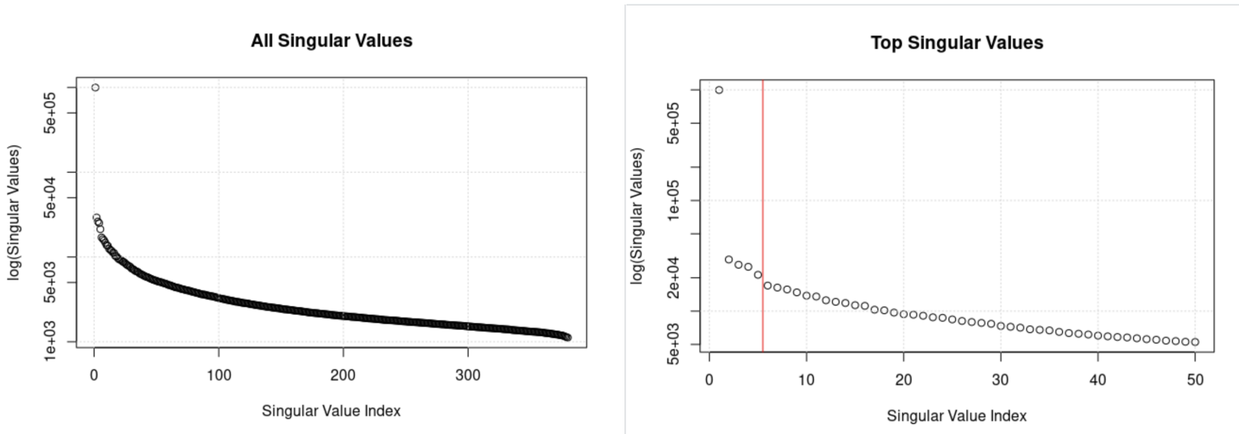
2.2 Singular Value Decomposition

We first perform singular value decomposition (SVD) on the unfolded full matrix $A_{f\times(np)}$. Our singular value decomposition can be denoted as:

$$\mathbf{A_{f\times(np)}} = \mathbf{U_{f\times f}}\mathbf{\Sigma_{f\times(np)}}\mathbf{V^{T}_{(np)\times(np)}}, \tag{1}$$

where Σ is a rank-f diagonal matrix with singular values on the diagonal. We subsequently organize the singular values in a decreasing order to determine the number of ranks that contain the most amount of information while keeping down the noise (Fig. 1.

From the semilog-plot of the singular values, we can see that other than the large drop after the first singular value, a reasonable cutoff of singular values would be after the fifth-largest singular value (Fig. 1).



(a) Semilog plot of all singular values (b) Semilog plot of top 50 singular values

Fig. 1: SVD for the full unfolded matrix

3 Results

3.1 Variance Analysis

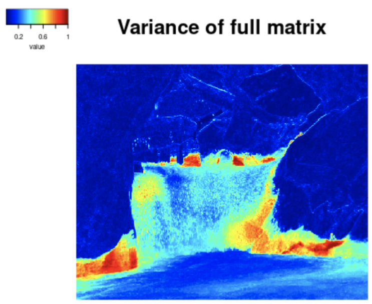
To visualize the elements in the matrix that are accounted for by the truncated SVDs, we plot the mean normalized variance in each pixel across all frames for each truncated SVD using 1 to 5 largest singular values. Here we included the rank-1 through rank-5 approximations, as well as the variance of pixels within the original matrix as a reference (Fig. 2)

From the plot of the full matrix, we are not surprised to see that the majority of the variance is located near the water and the splash zone, while the background and foreground rocks demonstrate relatively little variation. The slight variations on the rocks could potentially be slight changes in lighting during the period when the video is taken and are irrelevant to the waterfall motion, and can be considered as "noise". Interestingly, figures 2b to 2f did not recover the variation on the background rocks, which means that our rank-5 approximation was able to successfully de-noise the video.

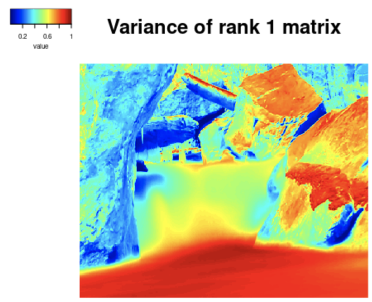
In the approximated matrices, notice that only the rank-1 approximation showed variability in all areas of the video. Starting with the rank-2 approximation, the variability near the "water splashes" far overpowers everywhere else in the video, including the waterfall itself. Some of the later singular vectors seem to recover some variability within the waterfall region, but the splashes still display a significantly larger variance in comparison.

3.2 Waterfall Decomposition

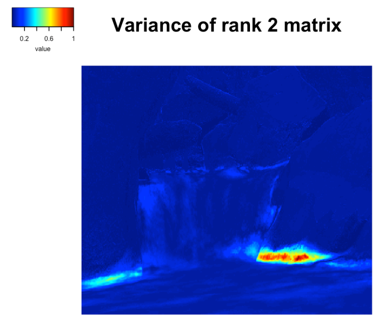
We visualized the first four components of the Singular Value Decomposition of the video. The first left singular vector shows the static background of the image (Fig 3a at top left), and the first right singular vector shows that it is constant (Fig 3b). The remaining singular vectors correspond to patterns in the water, which move in waves of about 30 frames, which is 1 second.



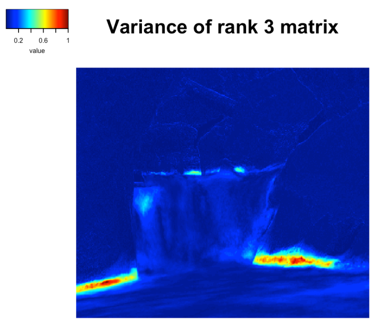
(a) original matrix



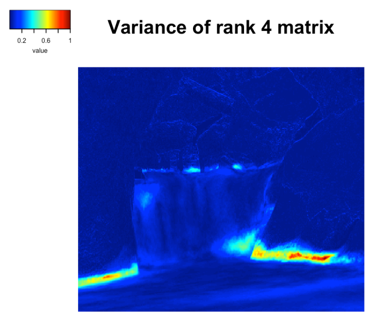
(b) rank 1 approximation



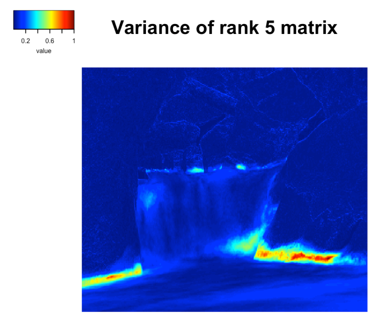
(c) rank 2 approximation



(d) rank 3 approximation

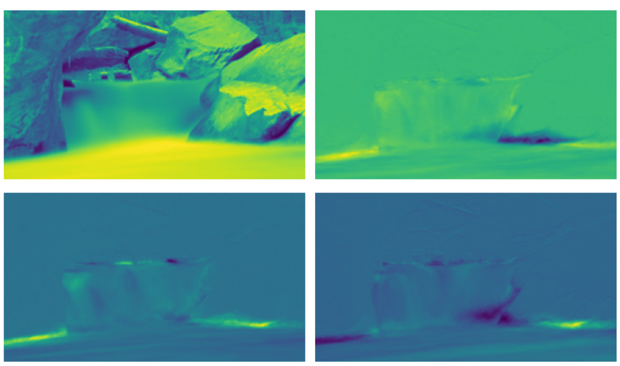


(e) rank 4 approximation

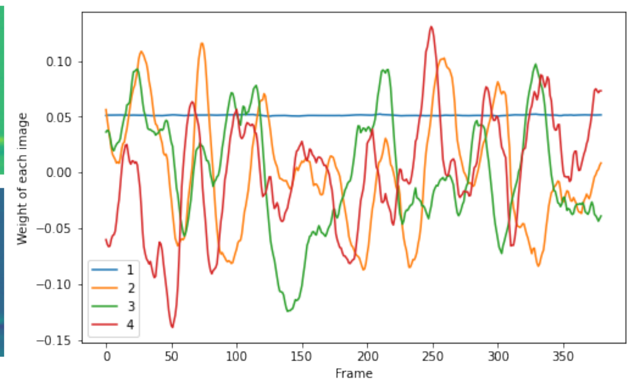


(f) rank 5 approximation

Fig. 2: Variance for the full unfolded matrix



(a) The first four left singular vectors, corresponding to the largest four singular values. These are vectors, but are reshaped to be shown as pictures here.



(b) The first four right singular values, corresponding to the largest four singular values.

Fig. 3: Top four components of the SVD of the full unfolded matrix

3.3 Reconstructing the Waterfall using rank-5 approximation

The reconstructed video according to the rank-5 approximation can be found here. We can recognize a clear distinction between the background, such as the rocks next to and behind the waterfall, and the waterfall itself. We notice that the background appears to be very detailed, while the majority of the details within the waterfall and the water splashes on the bottom are blurred out, even though they seem to have been picked up on the variance plots (Fig 2). Additionally, the reconstructed video does not display a clear directionality of the water flow – the waterfall only appears to change in brightness in patches.

4 Conclusion

We were able to apply SVD to a video and show that it separates the background from the foreground, and that the movement in the image is decomposed into gradual waves affecting regions of the image. We can see that the first 5 singular values capture much of the information in the video. The video reconstructed from those top 5 singular values looks like a de-noised version of the original video, though with the finer details dropped. In the reconstructed rank-5 video, the flow of time is no longer clear. This may be because the flow of water is usually shown through small details like moving water bubbles, which were dropped as noise.

We also see that the pixels in the background have lower variance than the pixels in the moving water. This difference provides an easy way to distinguish the foreground from the background.

4.1 Future Work

We would like to apply Dynamic Mode Decomposition (DMD) to this dataset. This would have two purposes. First, this method provides a visual way to learn and interpret DMD. Currently, it's difficult to see what is happening inside of a DMD analysis, such as which nodes in a network are affecting which others. With a video, it may be visually intuitive to see how information moves through the network of pixels. This applies to both simple DMD, and to DMD with control. And second, DMD may provide a way to create a perfectly-looped video, in which the final frame is identical to the first frame. Perfectly-looped videos are commonly used as backgrounds for other videos, and an automated method of creating them would be useful.

We would also like to separate parts of a video that move independently. To start, we could compute the correlation matrix between the pixels in this video, and then we could use the sign of the Fielder Vector.

5 Code Availability

The code of our analysis can be found at https://github.com/lwa19/waterfall.

References

[1] Hervé Abdi. "Singular value decomposition (SVD) and generalized singular value decomposition". In: *Encyclopedia of measurement and statistics* (2007), pp. 907–912.

[2] Rajesh A Jadav and Shailesh S Patel. "Application of singular value decomposition in image processing". In: *Indian journal of science and technology* 3.2 (2010), pp. 148–150.

[3] Cleve Moler. "Professor SVD". In: *TheMathWorksNews& Notes* (2006).

[4] Savage Interactive Pty Ltd. *Procreate*. Version 5.2.2. Nov. 1, 2021. URL: `https://procreate.art/`.

[5] Gilbert W Stewart. "On the early history of the singular value decomposition". In: *SIAM review* 35.4 (1993), pp. 551–566.

[6] Amit Surana, Geoff Patterson, and Indika Rajapakse. "Dynamic tensor time series modeling and analysis". In: *2016 IEEE 55th Conference on Decision and Control (CDC)*. IEEE. 2016, pp. 1637–1642.