Supplementary Material for "Deep Heterogeneous Hashing for Face Video Retrieval"

Shishi Qiao, *Student Member, IEEE,* Ruiping Wang, *Member, IEEE,* Shiguang Shan, *Senior Member, IEEE,* and Xilin Chen, *Fellow, IEEE*

In this document, we will give additional theoretical derivation of the corresponding sections in the main paper to support the method we proposed.

I. Derivation of the Variations

In this section, we will give the derivation of the two variations $d\mathbf{V}$ and $d\mathbf{\Sigma}$ introduced in Eqn.(11) of Sec.III.C of the main paper.

Let $\mathbf{D} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T$ by way of SVD, with $\mathbf{D} \in \mathbb{R}^{m\times n}$ and $m \le n$ (it is different from the case $m \ge n$ in [1]–[4], and thus different derivations of the variations), $\mathbf{\Sigma} \in \mathbb{R}^{m\times n}$ diagonal and $\mathbf{U} \in \mathbb{R}^{m\times m}$, $\mathbf{V} \in \mathbb{R}^{n\times n}$ orthogonal.

For the given variation $d\mathbf{D}$ of $\mathbf{D}$, our goal is to compute the variations $d\mathbf{\Sigma}, d\mathbf{V}$ in **two steps**. Following [1], we constrain that the derivation of the variation involves the invariants associated to its variables. Specifically, $d\mathbf{\Sigma}$ is diagonal, like $\mathbf{\Sigma}$, and $d\mathbf{U}$ and $d\mathbf{V}$ are orthogonal thus satisfy the constraints $\mathbf{U}^T d\mathbf{U} + d\mathbf{U}^T\mathbf{U} = \mathbf{0}$ and $\mathbf{V}^T d\mathbf{V} + d\mathbf{V}^T\mathbf{V} = \mathbf{0}$, respectively (It is trivial to infer that $\mathbf{U}^T d\mathbf{U}$ and $d\mathbf{V}^T\mathbf{V}$ are antisymmetric).

Step 1: Derivation of $d\mathbf{\Sigma}$. Firstly, we give the derivation of the variation $d\mathbf{\Sigma}$ with aforementioned conditions. Taking the variation of $\mathbf{D}$, we have

$$d\mathbf{D} = d\mathbf{U}\mathbf{\Sigma}\mathbf{V}^T + \mathbf{U}d\mathbf{\Sigma}\mathbf{V}^T + \mathbf{U}\mathbf{\Sigma}d\mathbf{V}^T \tag{S1}$$

since $\mathbf{U}$ and $\mathbf{V}$ are orthogonal, we have

$$\begin{aligned} &\mathbf{U}^T d\mathbf{D}\mathbf{V} = \mathbf{U}^T d\mathbf{U}\mathbf{\Sigma} + d\mathbf{\Sigma} + \mathbf{\Sigma}d\mathbf{V}^T\mathbf{V} \\ &\Rightarrow \mathbf{R} = \mathbf{A}\mathbf{\Sigma} + d\mathbf{\Sigma} + \mathbf{\Sigma}\mathbf{B} \end{aligned} \tag{S2}$$

where $\mathbf{R} = \mathbf{U}^T d\mathbf{D}\mathbf{V}$ and $\mathbf{A} = \mathbf{U}^T d\mathbf{U}$, $\mathbf{B} = d\mathbf{V}^T\mathbf{V}$. Since both $\mathbf{A}$ and $\mathbf{B}$ are antisymmetric, $\mathbf{A}\mathbf{\Sigma}$ and $\mathbf{\Sigma}\mathbf{B}$ have both zero diagonal. Furthermore, since $d\mathbf{\Sigma}$ is diagonal, we obtain $d\mathbf{\Sigma}$ as:

$$d\mathbf{\Sigma} = \mathbf{R}_{diag} \Rightarrow d\mathbf{\Sigma} = (\mathbf{U}^T d\mathbf{D}\mathbf{V})_{diag} \tag{S3}$$

where we denote $\mathbf{X}_{diag}$ an arbitrary matrix $\mathbf{X}$ with all off-diagonal elements being 0.

Step 2: Derivation of $d\mathbf{V}$. Since $m \le n$, the system of equations in [1] for solving the derivation of $d\mathbf{V}$ is undetermined in this case. Instead, we first construct a determined equation system to derive the variation $d\mathbf{U}$ in step 2.1, and further derive the final variation $d\mathbf{V}$ in step 2.2 with the conclusions of step 2.1, Eqn.(S2) and Eqn.(S3).

Step 2.1: From Eqn.(S2), another conclusion can be reached that

$$\begin{aligned} &\mathbf{A}\mathbf{\Sigma} + \mathbf{\Sigma}\mathbf{B} = \mathbf{R} - \mathbf{R}_{diag} \\ &\Rightarrow \mathbf{A}\mathbf{\Sigma}\mathbf{\Sigma}^T + \mathbf{\Sigma}\mathbf{B}\mathbf{\Sigma}^T = (\mathbf{R} - \mathbf{R}_{diag})\mathbf{\Sigma}^T = \bar{\mathbf{R}}\mathbf{\Sigma}^T \end{aligned} \tag{S4}$$

where, $\bar{\mathbf{R}} = \mathbf{R} - \mathbf{R}_{diag}$.

With $\mathbf{A}$ and $\mathbf{B}$ being both antisymmetric, one can construct the determined equation system and solve it as:

$$\begin{aligned} &\Rightarrow \begin{cases} \sigma_i \mathbf{B}_{ij}\sigma_j + \mathbf{A}_{ij}\sigma_j^2 = \bar{\mathbf{R}}_{ij}\sigma_j \\ -\sigma_j \mathbf{B}_{ij}\sigma_i - \mathbf{A}_{ij}\sigma_i^2 = \bar{\mathbf{R}}_{ji}\sigma_i \end{cases} \\ &\Rightarrow \mathbf{A}_{ij}(\sigma_j^2 - \sigma_i^2) = \bar{\mathbf{R}}_{ij}\sigma_j + \sigma_i\bar{\mathbf{R}}_{ji} \\ &\Rightarrow \mathbf{A}_{ij} = \begin{cases} (\sigma_j^2 - \sigma_i^2)^{-1}(\bar{\mathbf{R}}_{ij}\sigma_j + \sigma_i\bar{\mathbf{R}}_{ji}), & i \ne j \\ 0, & i = j \end{cases} \end{aligned} \tag{S5}$$

where $\sigma_i = \mathbf{\Sigma}_{ii}$. Rewrite the solution of $\mathbf{A}$ as $\mathbf{A} = \mathbf{P} \circ (\mathbf{\Sigma}\bar{\mathbf{R}}^T + \bar{\mathbf{R}}\mathbf{\Sigma}^T) = \mathbf{P} \circ (\mathbf{\Sigma}\mathbf{R}^T + \mathbf{R}\mathbf{\Sigma}^T)$, where

$$\mathbf{P}_{ij} = \begin{cases} \dfrac{1}{\sigma_j^2 - \sigma_i^2}, & i \ne j \\ 0, & i = j \end{cases} \tag{S6}$$

Furthermore, with the definition $\mathbf{A} = \mathbf{U}^T d\mathbf{U}$ in Eqn.(S2) and the orthogonality of $\mathbf{U}$, we obtain $d\mathbf{U}$ as:

$$d\mathbf{U} = \mathbf{U}\mathbf{A} \Rightarrow d\mathbf{U} = 2\mathbf{U}(\mathbf{P} \circ (\mathbf{U}^T d\mathbf{D}\mathbf{V}\mathbf{\Sigma}^T)_{sym}) \tag{S7}$$

where we denote $\circ$ the Hadamard product, $\mathbf{X}_{sym} = \frac{1}{2}(\mathbf{X} + \mathbf{X}^T)$. Note that this satisfies the condition $\mathbf{U}^T d\mathbf{U} + d\mathbf{U}^T\mathbf{U} = \mathbf{0}$, and thus preserves the orthogonality of $\mathbf{U}$.

Step 2.2: Using the $d\mathbf{\Sigma}$ and $d\mathbf{U}$ obtained, one can transform Eqn.(S1) to the form as:

$$\begin{aligned} &d\mathbf{D} = d\mathbf{U}\mathbf{\Sigma}\mathbf{V}^T + \mathbf{U}d\mathbf{\Sigma}\mathbf{V}^T + \mathbf{U}\mathbf{\Sigma}d\mathbf{V}^T \\ &\Rightarrow \mathbf{\Sigma}d\mathbf{V}^T = \mathbf{U}^T d\mathbf{D} - \mathbf{U}^T d\mathbf{U}\mathbf{\Sigma}\mathbf{V}^T - d\mathbf{\Sigma}\mathbf{V}^T := \mathbf{H} \end{aligned} \tag{S8}$$

This equation admits any solution of the block form $d\mathbf{V} = (d\mathbf{V}_1 \mid d\mathbf{V}_2)$, where $d\mathbf{V}_1^T := \mathbf{\Sigma}_m^{-1}\mathbf{H} \in \mathbb{R}^{m\times n}$ ($\mathbf{\Sigma}_m$ being the left m columns of $\mathbf{\Sigma}$) and $d\mathbf{V}_2^T \in \mathbb{R}^{(n-m)\times n}$ arbitrary. To determine $d\mathbf{V}_2^T$ uniquely, we resort to the orthogonality condition

$$\begin{aligned} &\mathbf{V}^T d\mathbf{V} + d\mathbf{V}^T\mathbf{V} = \mathbf{0} \\ &\Rightarrow \begin{pmatrix} \mathbf{V}_1^T d\mathbf{V}_1 + d\mathbf{V}_1^T\mathbf{V}_1 & \mathbf{V}_1^T d\mathbf{V}_2 + d\mathbf{V}_1^T\mathbf{V}_2 \\ \mathbf{V}_2^T d\mathbf{V}_1 + d\mathbf{V}_2^T\mathbf{V}_1 & \mathbf{V}_2^T d\mathbf{V}_2 + d\mathbf{V}_2^T\mathbf{V}_2 \end{pmatrix} = \mathbf{0} \end{aligned} \tag{S9}$$

Since that $\mathbf{V}_1^T\mathbf{V}_1 = \mathbf{I}$ and $\mathbf{V}_2^T\mathbf{V}_2 = \mathbf{I}$ by the orthogonality of $\mathbf{V}$, the block $d\mathbf{V}_1^T$ and $d\mathbf{V}_2^T$ already satisfy the top left equation and bottom right equation respectively, and we only need to check the bottom left (or the top right). We can verify that $d\mathbf{V}_2^T = -\mathbf{V}_2^T d\mathbf{V}_1\mathbf{V}_1^T$. As this also satisfies the remaining equation, orthogonality is satisfied. Finally, we obtain $d\mathbf{V}$ as:

$$d\mathbf{V} = (d\mathbf{V}^T)^T \Rightarrow d\mathbf{V} = (\mathbf{H}^T\mathbf{\Sigma}_m^{-1} \mid -\mathbf{V}_1\mathbf{\Sigma}_m^{-1}\mathbf{H}\mathbf{V}_2) \quad \text{(S10)}$$

With Eqn.(S3) and Eqn.(S10), we can replace $d\mathbf{\Sigma}$ and $d\mathbf{V}$ in Eqn.(11) of the main paper with their expressions w.r.t. $d\mathbf{D}$ to obtain the partial derivatives in Eqn.(13) of the main paper.

II. Basic identities

In this section we present some basic linear algebra identities of matrix inner product that are useful in the computations of matrix backpropagation. These identities are also presented in [1] and we list them here for convenient reference and better understanding of our derivations.

$$\mathbf{A} : \mathbf{B} = \mathbf{A}^T : \mathbf{B}^T = \mathbf{B} : \mathbf{A} \quad \text{(S11)}$$

$$\mathbf{A} : (\mathbf{B}\mathbf{C}) = (\mathbf{B}^T\mathbf{A}) : \mathbf{C} = (\mathbf{A}\mathbf{C}^T) : \mathbf{B} \quad \text{(S12)}$$

$$\mathbf{A} : \mathbf{B}_{diag} = \mathbf{A}_{diag} : \mathbf{B} \quad \text{(S13)}$$

$$\mathbf{A} : \mathbf{B}_{sym} = \mathbf{A}_{sym} : \mathbf{B} \quad \text{(S14)}$$

$$\mathbf{A} : (\mathbf{B} \circ \mathbf{C}) = (\mathbf{B} \circ \mathbf{A}) : \mathbf{C} \quad \text{(S15)}$$

$$(\mathbf{A}_1|\mathbf{A}_2) : (\mathbf{B}_1|\mathbf{B}_2) = \mathbf{A}_1 : \mathbf{B}_1 + \mathbf{A}_2 : \mathbf{B}_2 \quad \text{(S16)}$$

References

[1] C. Ionescu, O. Vantzos, and C. Sminchisescu, "Training deep networks with structured layers by matrix backpropagation," *CoRR*, vol. abs/1509.07838, 2015.
[2] Q. Wang, P. Li, and L. Zhang, "G2Denet: Global gaussian distribution embedding network and its application to visual recognition," in *IEEE,CVPR*, 2017, pp. 6507–6516.
[3] Z. Huang and L. J. V. Gool, "A riemannian network for SPD matrix learning," in *AAAI*, 2017, pp. 2036–2042.
[4] C. Ionescu, O. Vantzos, and C. Sminchisescu, "Matrix backpropagation for deep networks with structured layers," in *IEEE, ICCV*, 2015, pp. 2965–2973.