Supplementary material

Binshuai Wang[1], Xianglong Liu[1], Ke Xia[1], Kotagiri Ramamohanarao[2], and Dacheng Tao[3]

[1] State Key Lab of Software Development Environment, Beihang University, China
[2] Department of CIS, The University of Melbourne, Australia
[3] UBTECH Sydney AI Centre, School of IT, FEIT, University of Sydney, Australia

1 Proofs of the Theorems

This section is mainly to give the full proofs which is not shown all in the paper. $\Gamma(\cdot)$ denotes Gamma function.

A random variable $\mathbf{v}$, which is uniformly distributed in the unit sphere $\{\mathbf{v} \in \mathbb{R}^n | \, \|\mathbf{v}\|_2 = 1\}$, can be expressed by the following form:

$$\mathbf{v} = \frac{(X_1, X_2, ...X_n)}{\sqrt{\Sigma_{i=1}^n X_i^2}}, X_i \sim \mathcal{N}(0,1) \tag{1}$$

where X_i is a random variable which has a standard normal distribution.

Theorem 1 (Special case) *Given subspaces* S_1 *and* S_2 *in* $\mathbb{R}^n$ *with dimension* $d_1 = d_2 = 1$,

$$\alpha_0 = \frac{\sqrt{2}}{\sqrt{n^3 + 2n^2}} - \frac{1}{n}, \tag{2}$$

if $\mathbf{v}$ *is a random variable uniformly distributed in the unit sphere, then we have the following equation*

$$\mathbf{E}\left[(z_{\mathbf{v}}(\mathrm{S}_1) z_{\mathbf{v}}(\mathrm{S}_2)\right] = \frac{2}{(n+2)n} \left\|\mathbf{P}_1^\top \mathbf{P}_2\right\|_F^2 \tag{3}$$

where $\mathbf{P}_1$ *and* $\mathbf{P}_2$ *are any orthonormal basis of subspace* S_1 *and* S_2*.*

Proof Because of the uniform distribution in the direction of the random variable $\mathbf{v}$, without loss of generality, it can be supposed that

$$\mathbf{P}_1 = (1, 0, ...0)^\top, \mathbf{P}_2 = (u_1, u_2, 0...0)^\top \tag{4}$$

$$u_1 = \mathbf{P}_1 \cdot \mathbf{P}_2, u_2 = \sqrt{1 - u_1^2} \tag{5}$$

Firstly, we will compute $\mathbf{E}[\alpha_{\mathbf{v}}(\mathrm{S}_1)\alpha_{\mathbf{v}}(\mathrm{S}_2)]$, $\mathbf{E}[\alpha_{\mathbf{v}}(\mathrm{S}_1)]$, $\mathbf{E}[\alpha_{\mathbf{v}}(\mathrm{S}_2)]$.

Because

$$\begin{aligned} \alpha_{\mathbf{v}}(\mathrm{S}_1) &= \frac{X_1^2}{\sum_{i=1}^n X_i^2} \\ \alpha_{\mathbf{v}}(\mathrm{S}_2) &= \frac{(u_1 X_1 + u_2 X_2)^2}{\sum_{i=1}^n X_i^2} \end{aligned} \tag{6}$$

we could get

$$\mathbf{E}[\alpha_{\mathbf{v}}(\mathrm{S}_1)\alpha_{\mathbf{v}}(\mathrm{S}_2)] = \mathbf{E}\left(\frac{X_1^2}{\sum_{i=1}^n X_i^2}\cdot\frac{(u_1X_1+u_2X_2)^2}{\sum_{i=1}^n X_i^2}\right)$$

$$= \int_{-\infty}^{+\infty} ... \int_{-\infty}^{+\infty} \frac{x_1^2}{\sum_{i=1}^n x_i^2}\cdot\frac{(u_1x_1+u_2x_2)^2}{\sum_{i=1}^n x_i^2}\cdot\prod_{i=1}^n \frac{e^{-\frac{x_i^2}{2}}}{\sqrt{2\pi}} dx_1...dx_n \tag{7}$$

Based on variable transformation, we get

$$\begin{aligned} x_1 &= r\cos(t_1) \\ x_2 &= r\sin(t_1)\cos(t_2) \\ &... \\ x_{n-1} &= r\sin(t_1)\sin(t_2)...\sin(t_{n-2})\cos(t_{n-1}) \\ x_n &= r\sin(t_1)\sin(t_2)...\sin(t_{n-2})\sin(t_{n-1}) \end{aligned} \tag{8}$$

in which,

$$\begin{aligned} &0 \le r < +\infty \\ &0 \le t_1 \le \pi, ..., 0 \le t_{n-2} \le \pi, 0 \le t_{n-1} \le 2\pi \\ \text{subject to}\quad & \\ &x_1^2 + x_2^2 + ... + x_n^2 = r^2 \end{aligned} \tag{9}$$

and

$$\begin{aligned} J &= \frac{\partial(x_1, x_2, ...x_n)}{\partial(r, t_1, t_2, ...t_{n-1})} \\ &= r^{n-1}\sin^{n-2}t_1\sin^{n-3}t_2...\sin t_{n-2} \end{aligned} \tag{10}$$

Because,

$$\frac{x_1^2}{\sum_{i=1}^n x_i^2} = \frac{r^2\cos^2 t_1}{r^2} = \cos^2 t_1 \tag{11}$$

$$\frac{(u_1x_1+u_2x_2)^2}{\sum_{i=1}^n x_i^2} = (u_1\cos t_1 + u_2\sin t_1\cos t_2)^2 \tag{12}$$

$$\prod_{i=1}^n \frac{e^{-\frac{x_i^2}{2}}}{\sqrt{2\pi}} = \frac{e^{-\frac{1}{2}\sum_{i=1}^n x_i^2}}{\sqrt{2\pi}^n} = \frac{e^{-\frac{1}{2}r^2}}{\sqrt{2\pi}^n} \tag{13}$$

So, it follows that

$$\int_{-\infty}^{+\infty} ... \int_{-\infty}^{+\infty} \frac{x_1^2}{\sum_{i=1}^n x_i^2}\cdot\frac{(u_1x_1+u_2x_2)^2}{\sum_{i=1}^n x_i^2}\cdot\prod_{i=1}^n \frac{e^{-\frac{x_i^2}{2}}}{\sqrt{2\pi}} dx_1...dx_n$$

$$= \int_0^{\pi} ... \int_0^{+\infty} \int_0^{2\pi} \cos^2 t_1 (u_1\cos t_1 + u_2\sin t_1\cos t_2)^2 \frac{e^{-\frac{1}{2}r^2}}{\sqrt{2\pi}^n} J dt_1...drdt_{n-1} \tag{14}$$

Denoting

$$
\begin{aligned}
K_1 &= \int_0^{+\infty} e^{-\frac{r^2}{2}} r^{n-1}\, dr \\
K_2 &= \int_0^{\pi}\int_0^{\pi} \dots \int_0^{\pi} \sin^{n-4} t_3 \sin^{n-5} t_4 \dots \sin t_{n-2} dt_3 dt_4 \dots dt_{n-2} \\
K_3 &= \int_0^{\pi}\int_0^{\pi} \cos^2 t_1 (u_1 \cos t_1 + u_2 \sin t_1 \cos t_2)^2 \sin^{n-2} t_1 \sin^{n-3} t_2 dt_1 dt_2 \\
K_4 &= \int_0^{2\pi} 1\, dt_{n-1} \\
K_5 &= \frac{1}{\sqrt{2\pi}^n}
\end{aligned}
\tag{15}
$$

Now, we will separately compute K_1, K_2, K_3.

(a)Computing K_1

$$\int_0^{+\infty} e^{-\frac{r^2}{2}} r^{n-1}\, dr \tag{16}$$

Using the proposition of Gamma function, we can get

$$\int_0^{+\infty} e^{-\frac{r^2}{2}} r^{n-1}\, dr = \begin{cases} (m-1)!2^{m-1} & n = 2m \\ (2m-1)!!\sqrt{\dfrac{\pi}{2}} & n = 2m+1 \end{cases} \tag{17}$$

(b)Computing K_2

Because

$$\int_0^{\pi} \sin^n(t)\mathrm{dt} = 2\int_0^{\frac{\pi}{2}} \sin^n(t)\mathrm{dt} = \begin{cases} \frac{(2m-1)!!}{(2m)!!}\cdot\pi = \frac{(n-1)!!}{n!!}\cdot\pi & n = 2m \\ \frac{(2m)!!}{(2m+1)!!}\cdot 2 = \frac{(n-1)!!}{n!!}\cdot 2 & n = 2m+1 \end{cases} \tag{18}$$

So,

$$K_2 = \begin{cases} \frac{2\pi^{m-1}}{(m-2)!} & n = 2m \\ \frac{2^m \pi^{m-1}}{(2m-3)!!} & n = 2m+1 \end{cases} \tag{19}$$

(c)Computing K_3

$$\int_0^{\pi} \left(\sin^4(t_1) - 2\sin^2(t_1) + 1\right) \sin^{n-2} t_1\, dt_1 = \begin{cases} \pi\frac{(n-3)!!}{(n+2)!!}\cdot 3 & n = 2m \\ \frac{(n-3)!!}{(n+2)!!}\cdot 6 & n = 2m+1 \end{cases} \tag{20}$$

$$\int_0^{\pi} \sin^{n-3}(t_2)\, dt_2 = \begin{cases} \frac{(n-4)!!}{(n-3)!!}\cdot 2 & n = 2m \\ \frac{(n-4)!!}{(n-3)!!}\cdot\pi & n = 2m+1 \end{cases} \tag{21}$$

$$\int_0^{\pi} \left(1 - \sin^2 t_1\right) \sin^n t_1\, dt_1$$

$$=\begin{cases}\left(\frac{(n-1)!!}{n!!}-\frac{(n+1)!!}{(n+2)!!}\right)\cdot\pi=\frac{(n-1)!!}{(n+2)!!}\cdot\pi & n=2m\\ \left(\frac{(n-1)!!}{n!!}-\frac{(n+1)!!}{(n+2)!!}\right)\cdot 2=\frac{(n-1)!!}{(n+2)!!}\cdot 2 & n=2m+1\end{cases} \tag{22}$$

$$\int_0^{\pi}\left(1-\sin^2(t_2)\right)\sin^{n-3}(t_2)\,dt_2$$

$$=\begin{cases}\left(\frac{(n-4)!!}{(n-3)!!}-\frac{(n-2)!!}{(n-1)!!}\right)\cdot 2=\frac{(n-4)!!}{(n-1)!!}\cdot 2 & n=2m\\ \left(\frac{(n-4)!!}{(n-3)!!}-\frac{(n-2)!!}{(n-1)!!}\right)\cdot\pi=\frac{(n-4)!!}{(n-1)!!}\cdot\pi & n=2m+1\end{cases} \tag{23}$$

$$\int_0^{\pi}\int_0^{\pi}u_1^2\cos^4(t_1)\sin^{n-2}(t_1)\sin^{n-3}(t_2)\,dt_1dt_2$$

$$=u_1^2\int_0^{\pi}\left(\sin^4(t_1)-2\sin^2(t_1)+1\right)\sin^{n-2}(t_1)\,dt_1\cdot\int_0^{\pi}\sin^{n-3}(t_2)\;dt_2$$

$$=\begin{cases}\pi u_1^2\frac{(n-3)!!}{(n+2)!!}\cdot 3\frac{(n-4)!!}{(n-3)!!}\cdot 2 & n=2m\\ u_1^2\frac{(n-3)!!}{(n+2)!!}\cdot 6\frac{(n-4)!!}{(n-3)!!}\cdot\pi & n=2m+1\end{cases}$$

$$=\begin{cases}\pi u_1^2\frac{1}{(n+2)n(n-2)}\cdot 6 & n=2m\\ \pi u_1^2\frac{1}{(n+2)n(n-2)}\cdot 6 & n=2m+1\end{cases}=\frac{6\pi u_1^2}{(n+2)n(n-2)} \tag{24}$$

$$\int_0^{\pi}\int_0^{\pi}u_2^2\cos^2 t_1\cos^2 t_2\sin^2 t_1\sin^{n-2}t_1\sin^{n-3}t_2\,dt_1dt_2$$

$$\begin{aligned}&=u_2^2\int_0^{\pi}\left(1-\sin^2(t_1)\right)\sin^{n}(t_1)\;dt_1\cdot\int_0^{\pi}\left(1-\sin^2(t_2)\right)\sin^{n-3}(t_2)\;dt_2\\&=\begin{cases}u_2^2\frac{(n-1)!!}{(n+2)!!}\cdot\pi\frac{(n-4)!!}{(n-1)!!}\cdot 2 & n=2m\\ u_2^2\frac{(n-1)!!}{(n+2)!!}\cdot 2\frac{(n-4)!!}{(n-1)!!}\cdot\pi & n=2m+1\end{cases}\\&=\begin{cases}\frac{1}{(n+2)n(n-2)}\cdot 2\pi u_2^2 & n=2m\\ \frac{1}{(n+2)n(n-2)}\cdot 2\pi u_2^2 & n=2m+1\end{cases}=\frac{2\pi u_2^2}{(n+2)n(n-2)}\end{aligned} \tag{25}$$

$$\int_0^{\pi}\int_0^{\pi}2u_1u_2\cos^3(t_1)\sin(t_1)\cos(t_2)\sin^{n-2}(t_1)\sin^{n-3}(t_2)\,dt_1dt_2=0 \tag{26}$$

We get

$$K_3=\frac{6\pi u_1^2}{(n+2)n(n-2)}+\frac{2\pi u_2^2}{(n+2)n(n-2)}=\frac{4\pi u_1^2+2\pi}{(n+2)n(n-2)} \tag{27}$$

Therefore, it follows that

$$\mathbf{E}[\alpha_{\mathbf{v}}(\mathrm{S}_1)\alpha_{\mathbf{v}}(\mathrm{S}_2)]=K_1K_2K_3K_4K_5=\frac{2(u_1)^2}{(n+2)n}+\frac{1}{(n+2)n} \tag{28}$$

And because

$$\mathbf{E}[\alpha_{\mathbf{v}}(\mathrm{S}_1)]=\frac{X_1^2}{\sum_{i=1}^{n}X_i^2} \tag{29}$$

and because of the symmetry

$$n\mathbf{E}[\alpha_{\mathbf{v}}(\mathrm{S}_1)] = \frac{nX_1^2}{\sum_{i=1}^{n} X_i^2} = 1 \tag{30}$$

we can get

$$\mathbf{E}[\alpha_{\mathbf{v}}(\mathrm{S}_1)] = \mathbf{E}[\alpha_{\mathbf{v}}(\mathrm{S}_2)] = \frac{1}{n} \tag{31}$$

Secondly, we will compute $\mathbf{E}\left[(z_{\mathbf{v}}(\mathrm{S}_1)z_{\mathbf{v}}(\mathrm{S}_2)\right]$.

By bringing an undetermined α_0, and expand $\mathbf{E}\left[(z_{\mathbf{v}}(\mathrm{S}_1)z_{\mathbf{v}}(\mathrm{S}_2)\right]$ by (28), (31), we can get

$$\mathbf{E}\left[(z_{\mathbf{v}}(\mathrm{S}_1)z_{\mathbf{v}}(\mathrm{S}_2)\right] = \frac{2u_1^2}{(n+2)n} + \alpha_0^2 + \frac{2\alpha_0}{n} + \frac{1}{(n+2)n} \tag{32}$$

And let

$$\alpha_0 = -\frac{\sqrt{2}}{\sqrt{n^3+2n^2}} - \frac{1}{n} \text{or } \frac{\sqrt{2}}{\sqrt{n^3+2n^2}} - \frac{1}{n} \tag{33}$$

such that

$$\alpha_0^2 + \frac{2\alpha_0}{n} + \frac{1}{(n+2)n} = 0 \tag{34}$$

Therefore we can infer that

$$\mathbf{E}\left[(z_{\mathbf{v}}(\mathrm{S}_1)z_{\mathbf{v}}(\mathrm{S}_2)\right] = \frac{2u_1^2}{(n+2)n} = \frac{2(\mathbf{P}_1 \cdot \mathbf{P}_2)^2}{(n+2)n} = \frac{2\|\mathbf{P}_1^\top \mathbf{P}_2\|_F^2}{(n+2)n} \tag{35}$$

Theorem 1 (General case) *Given subspaces* S_1 *and* S_2 *in* $\mathbb{R}^n$*, with dimension* d_1 *and* d_2 *respectively and*

$$\alpha_0 = \frac{\sqrt{2}}{\sqrt{n^3+2n^2}} - \frac{1}{n}, \tag{36}$$

if $\mathbf{v}$ *is a random variable uniformly distributed in the unit sphere, then we have the following equation*

$$\mathbf{E}\left[(z_{\mathbf{v}}(\mathrm{S}_1)z_{\mathbf{v}}(\mathrm{S}_2)\right] = \frac{2}{(n+2)n}\left\|\mathbf{P}_1^\top \mathbf{P}_2\right\|_F^2 \tag{37}$$

where $\mathbf{P}_1$ *and* $\mathbf{P}_2$ *are any orthonormal basis of subspace* S_1 *and* S_2*.*

Proof Writing $\mathbf{P}_1$, $\mathbf{P}_2$ as following forms.

$$\mathbf{P}_1 = (\mathbf{a}_1, \mathbf{a}_2, ...\mathbf{a}_{d_1}), \mathbf{P}_2 = (\mathbf{b}_1, \mathbf{b}_2, ...\mathbf{b}_{d_2}) \tag{38}$$

and denoting the 1-dimension subspace spanned by $\mathbf{a}_i, \mathbf{b}j$ as $\mathrm{A}_i, \mathrm{B}_j$. It is clear to see that

$$\begin{aligned} z_{\mathbf{v}}(\mathrm{S}_1) &= \|\mathbf{P}_1^\top \mathbf{v}\|_F^2 + d_1\alpha_0 = \Sigma_{i=1}^{d_1}(\|\mathbf{a}_i^\top \mathbf{v}\|_F^2 + \alpha_0) = \Sigma_{i=1}^{d_1} z_{\mathbf{v}}(\mathrm{A}_i) \\ z_{\mathbf{v}}(\mathrm{S}_2) &= \|\mathbf{P}_2^\top \mathbf{v}\|_F^2 + d_2\alpha_0 = \Sigma_{j=1}^{d_2}(\|\mathbf{b}_j^\top \mathbf{v}\|_F^2 + \alpha_0) = \Sigma_{j=1}^{d_2} z_{\mathbf{v}}(\mathrm{B}_j) \end{aligned} \tag{39}$$

So,

$$\begin{aligned}&\mathbf{E}\left[(z_{\mathbf{v}}(\mathrm{S}_1)z_{\mathbf{v}}(\mathrm{S}_2)\right]\\=&\mathbf{E}(\Sigma_{i=1}^{d_1}z_{\mathbf{v}}(\mathrm{A}_i)\Sigma_{j=1}^{d_2}z_{\mathbf{v}}(\mathrm{B}_j))\\=&\mathbf{E}(\Sigma_{i=1}^{d_1}\Sigma_{j=1}^{d_2}(z_{\mathbf{v}}(\mathrm{A}_i)z_{\mathbf{v}}(\mathrm{B}_j)))\\=&\Sigma_{i=1}^{d_1}\Sigma_{j=1}^{d_2}\mathbf{E}(z_{\mathbf{v}}(\mathrm{A}_i)z_{\mathbf{v}}(\mathrm{B}_j))\end{aligned}\tag{40}$$

According to (3),

$$\begin{aligned}&\Sigma_{i=1}^{d_1}\Sigma_{j=1}^{d_2}\mathbf{E}(z_{\mathbf{v}}(\mathrm{A}_i)z_{\mathbf{v}}(\mathrm{B}_j))\\=&\Sigma_{i=1}^{d_1}\Sigma_{j=1}^{d_2}\frac{2}{(n+2)n}\|\mathbf{a}_i^\top\mathbf{b}_j\|_F^2\\=&\frac{2}{(n+2)n}\Sigma_{i=1}^{d_1}\Sigma_{j=1}^{d_2}\|\mathbf{a}_i^\top\mathbf{b}_j\|_F^2\\=&\frac{2}{(n+2)n}\|\mathbf{P}_1^\top\mathbf{P}_2\|_F^2\end{aligned}\tag{41}$$

Theorem 2 *When m is large enough, $\mathcal{H}=f_{\mathbf{r}}\circ\mathcal{Z}$ is a locality sensitive hash defined on $(\Omega, dist)$ approximately, i.e., $\mathcal{H}$ is $(\delta,\delta(1+\epsilon),1-\delta,1-\delta(1+\epsilon))$-sensitive with $\delta,\epsilon>0$.*

Proof According to Theorem 1, for any two subspace S_1 and S_2 in $\mathbb{R}^n$, with dimension d_1 and d_2 ($d_1\le d_2$) respectively, we have

$$\begin{aligned}\lim_{m\to+\infty}\frac{1}{m}\mathbf{z}_{\mathrm{S}_1}^\top\mathbf{z}_{\mathrm{S}_2}&=\frac{2}{n(n+2)}\sum_{i=1}^{d_1}\cos^2(\xi_i)\,,\\\lim_{m\to+\infty}\frac{1}{m}\|\mathbf{z}_{\mathrm{S}_1}\|_2^2&=\frac{2}{n(n+2)}d_1,\\\lim_{m\to+\infty}\frac{1}{m}\|\mathbf{z}_{\mathrm{S}_2}\|_2^2&=\frac{2}{n(n+2)}d_2.\end{aligned}$$

Therefore,

$$\lim_{m\to+\infty}\frac{\mathbf{z}_{\mathrm{S}_1}^\top\mathbf{z}_{\mathrm{S}_2}}{\|\mathbf{z}_{\mathrm{S}_1}\|_2\|\mathbf{z}_{\mathrm{S}_2}\|_2}=\frac{\sum_{i=1}^{d_1}\cos^2(\xi_i)}{\sqrt{d_1}\sqrt{d_2}},$$

when m is large enough, we approximately have

$$\frac{\mathbf{z}_{\mathrm{S}_1}^\top\mathbf{z}_{\mathrm{S}_2}}{\|\mathbf{z}_{\mathrm{S}_1}\|_2\|\mathbf{z}_{\mathrm{S}_2}\|_2}\approx\frac{\sum_{i=1}^{d_1}\cos^2(\xi_i)}{\sqrt{d_1}\sqrt{d_2}}$$

and

$$\arccos\frac{\mathbf{z}_{\mathrm{S}_1}^\top\mathbf{z}_{\mathrm{S}_2}}{\|\mathbf{z}_{\mathrm{S}_1}\|_2\|\mathbf{z}_{\mathrm{S}_2}\|_2}\approx\arccos\frac{\sum_{i=1}^{d_1}\cos^2(\xi_i)}{\sqrt{d_1}\sqrt{d_2}}.$$

Therefore, the following fact holds,

$$
\begin{aligned}
&\Pr[f_{\mathbf{r}} \circ \mathcal{Z}(\mathrm{S}_1) = f_{\mathbf{r}} \circ \mathcal{Z}(\mathrm{S}_2)] \\
&= 1 - \frac{1}{\pi}\arccos(\frac{\mathbf{z}_{\mathrm{S}_1} \cdot \mathbf{z}_{\mathrm{S}_2}}{\|\mathbf{z}_{\mathrm{S}_1}\|_2 \|\mathbf{z}_{\mathrm{S}_2}\|_2}) \\
&\approx 1 - \frac{1}{\pi}\arccos\frac{\sum_{i=1}^{d_1} \cos^2(\xi_i)}{\sqrt{d_1}\sqrt{d_2}} \\
&= 1 - \mathrm{dist}(\mathrm{S}_1, \mathrm{S}_2)
\end{aligned}
$$

If $\mathrm{dist}(\mathrm{S}_1, \mathrm{S}_2) \leq \delta$, then we have

$$\Pr[f_{\mathbf{r}} \circ \mathcal{Z}(\mathrm{S}_1) = f_{\mathbf{r}} \circ \mathcal{Z}(\mathrm{S}_2)] \geq 1 - \delta = p_1$$

Likewise, when $\mathrm{dist}(\mathrm{S}_1, \mathrm{S}_2) > \delta(1+\epsilon)$, we have

$$\Pr[f_{\mathbf{r}} \circ \mathcal{Z}(\mathrm{S}_1) = f_{\mathbf{r}} \circ \mathcal{Z}(\mathrm{S}_2)] < 1 - \delta(1+\epsilon) = p_2,$$

where $p_2 < p_1$. This completes the proof.

2 Illustrative Examples

Example 1 In $\mathbb{R}^3$, let S_1 and S_2 denote the subspaces xOy and yOz of $\mathbb{R}^3$ and let $\mathbf{P}_1$ and $\mathbf{P}_2$ be the orthonormal matrice of subspace S_1, $\mathbf{P}_3$ be the orthonormal matrix of subspace S_2. $\mathbf{v} = \left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}, 0\right)^\top$, $\mathbf{P}_1 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{pmatrix}$,$\mathbf{P}_2 = \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ 0 & 0 \end{pmatrix}$,$\mathbf{P}_3 = \begin{pmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 1 \end{pmatrix}$.

$$
\begin{aligned}
\alpha_{\mathbf{v}}(\mathrm{S}_1) &= \left\|\mathbf{P}_1^\top \cdot \mathbf{v}\right\|_F^2 = \left\|\mathbf{P}_2^\top \cdot \mathbf{v}\right\|_F^2 = 1 \\
\alpha_{\mathbf{v}}(\mathrm{S}_2) &= \left\|\mathbf{P}_3^\top \cdot \mathbf{v}\right\|_F^2 = \frac{1}{2}
\end{aligned} \tag{42}
$$

$$\left\|\mathbf{P}_1^\top \cdot \mathbf{P}_3\right\|_F^2 = \left\|\mathbf{P}_2^\top \cdot \mathbf{P}_3\right\|_F^2 = 1 \tag{43}$$

$$\xi_1 = 0, \xi_2 = \pi/2 \tag{44}$$

Example 2 In $\mathbb{R}^2$, let S_1 and S_2 denote the subspaces of $\mathbb{R}^2$ and let $\mathbf{P}_1$ be the orthonormal matrix of subspace S_1, $\mathbf{P}_2$ be the orthonormal matrix of subspace S_2. $\mathbf{P}_1 = (1,0)^\top$, $\mathbf{P}_2 = (\cos\gamma, \sin\gamma)^\top$($\gamma$ is constant), $\mathbf{v} = (\cos\tau, \sin\tau)^\top, \tau \sim U[0, 2\pi]$

$$
\begin{aligned}
&\mathbf{E}[\alpha_{\mathbf{v}}(\mathrm{S}_1)\alpha_{\mathbf{v}}(\mathrm{S}_2)] \\
&= \frac{1}{2\pi}\int_0^{2\pi} ((1,0) \cdot (\cos\tau, \sin\tau)^\top)^2 ((\cos\gamma, \sin\gamma) \cdot (\cos\tau, \sin\tau)^\top)^2 \, d\tau
\end{aligned}
$$

$$\frac{\cos^2(\gamma)}{4} + \frac{1}{8} \tag{45}$$

$$\mathbf{E}[\alpha_{\mathbf{v}}(\mathrm{S}_2)] = \frac{1}{2\pi}\int_0^{2\pi} ((\cos\gamma, \sin\gamma).(\cos\tau, \sin\tau)^\top)^2 \, d\tau = \frac{1}{2} \tag{46}$$

$$\mathbf{E}[\alpha_{\mathbf{v}}(\mathrm{S}_1)] = \frac{1}{2\pi}\int_0^{2\pi} ((1,0).(\cos\tau, \sin\tau)^\top)^2 \, d\tau = \frac{1}{2} \tag{47}$$

Therefore,

$$\mathbf{E}[z_{\mathbf{v}}(\mathrm{S}_1)z_{\mathbf{v}}(\mathrm{S}_2)] = \mathbf{E}[\alpha_{\mathbf{v}}(\mathrm{S}_1)\alpha_{\mathbf{v}}(\mathrm{S}_2)] + \alpha_0\mathbf{E}[\alpha_{\mathbf{v}}(\mathrm{S}_1)] + \alpha_0\mathbf{E}[\alpha_{\mathbf{v}}(\mathrm{S}_2)] + \alpha_0^2 \tag{48}$$

Let $\alpha_0 = \frac{\sqrt{2}-2}{4}$

$$\mathbf{E}[z_{\mathbf{v}}(\mathrm{S}_1)z_{\mathbf{v}}(\mathrm{S}_2)] = \frac{\cos^2(\gamma)}{4} \tag{49}$$

3 More Experimental Results

3.1 Video Deduplication

In this experiments, we construct a subset of UQ_VIDEO [3] to show the scalability of the proposed RAP method.

Experiment Setup: In this experiments, we first construct the database set with more than 10,000 videos, rather than the whole database, and further remove those with very few frames. The database mainly consists of two parts: all the videos (more than 3,000) belonging to the 24 classes corresponding to the given 24 queries, and the randomly selected background videos. For each video, we use every $d = 5$ consecutive frames without overlap to form a subspace. At the search stage, we use the Hamming distance ranking strategy to rank all the video clips in the database set according to the averaged Hamming distances to the query.

Results: Table 1 shows the mean average precision (MAP) and search time using Hamming distance ranking in the video search task. We obtain the similar results to that on the whole dataset. Namely, the subspace hashing methods still perform much better than the state-of-the-art vector hashing methods. Besides, when using different number of hash bits, our RAP method can consistently achieve the best performance, and more than 8× faster than the random vector hashing method LSH.

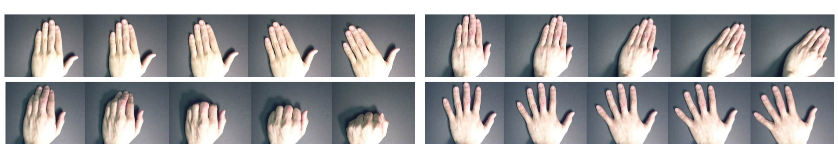

Fig. 1. Cambridge Gestures database

Table 1. Video retrieval performance on UQ_VIDEO.

Methods	MAP (%), $d = d_q = 5$			Search Time (s)		
	$b = 64$	$b = 128$	$b = 256$	$b = 64$	$b = 128$	$b = 256$
LSH	53.89 ± 2.75	55.71 ± 2.93	55.76 ± 0.29	0.3846	0.5213	0.7405
ITQ	49.58 ± 1.16	50.68 ± 0.94	-	0.3865	0.5178	-
DPLM	54.28 ± 1.28	61.76 ± 1.46	64.38 ± 1.03	0.3852	0.5268	0.7158
ABQ	36.18 ± 3.57	36.26 ± 1.35	38.74 ± 0.29	0.5217	0.7404	1.3365
BSS	53.28 ± 1.88	65.31 ± 0.70	76.31 ± 1.12	0.0685	0.0685	0.0779
RAP	**54.92** ± 2.99	**68.09** ± 2.15	**77.50** ± 1.18	0.0721	0.0753	0.0892

Table 2. Gesture recognition performance on Cambridge Gestures.

Methods	MAP (%), $b = 4 \times 10^3, m = 10^4$	Search Time (s)
GLH	95.00 ± 0.30	0.2850
BSS	93.17 ± 0.93	0.1273
RAP	95.22 ± 1.27	0.0345

3.2 Gesture Recognition

Gesture recognition involves the discrimination based on a number of successive similar images. Therefore, similar to the fact recognition, each set of consecutive images can be represented by a subspace, and the recognition can be completed based on the nearest subspace search.

Dataset: In this experiment, we employ the Cambridge Gestures database [2], containing 900 videos of hand gestures. There are three kinds of hand shapes: flat, spread, v-shape, and three kinds of motions: left, right, contraction. Therefore, the combination of hand shapes and motions yield 9 classes of gestures. Each class contains 100 samples, evenly divided into five sets with 20 samples each. We use one set as testing and the remaining four sets for the database. For each video sample, the middle 32 frames are selected to represent the video, each of which is transformed to gray-scale and rescaled to 20×20.

Experiment Setup: We follow the same strategy used in [1]. Namely, we first unfold the $20 \times 20 \times 32$ data cube along any of the three axes (x, y, t) to form three subspaces respectively. Then, three binary signatures with the same length are generated for the corresponding three subspaces. Finally, we concatenate the binary codes to form a single one for each data cube for the database and the testing set.

Results: Table 2 reports the precision performance of top-1 results and the search time. In this experiment, RAP gets close (even better) performance to the state-of-the-art subspace hashing methods GLH and BSS. Moreover, we can see that RAP can significantly boosts the online search process, compared to BSS and GLH.

References

1. Ji, J., Li, J., Tian, Q., Yan, S., Zhang, B.: Angular-similarity-preserving binary signatures for linear subspaces. IEEE TIP 24(11), 4372–4380 (2015)
2. Kim, T.K., Wong, S.F., Cipolla, R.: Tensor canonical correlation analysis for action classification. In: IEEE CVPR. pp. 1–8. IEEE (2007)
3. Song, J., Yang, Y., Huang, Z., Shen, H.T., Hong, R.: Multiple feature hashing for real-time large scale near-duplicate video retrieval. In: ACM MM. pp. 423–432. ACM (2011)