Extra Material

1. Data Preprocessing

1.1. Removing the DC Component with an Orthogonal Projection. The projector P_{remDC} is computed such that the first (for each color channel) component of $P_{remDC}\mathbf{x}$ corresponds to the DC component(s) of that patch. The transpose of the matrix

$$P = \begin{pmatrix} 1 & 0 & 0 & \cdots \\ 1 & 1 & 0 & \cdots \\ 1 & 0 & \ddots & \cdots \\ \vdots & & & 1 \end{pmatrix}$$

has exactly the required property. However, it is not an orthogonal transformation. Therefore, we decompose P into $P = QR$ where R is upper triangular and Q is an orthogonal transform. Since $P = QR$, the first column of Q must be a multiple of the vector with all coefficients equal to one (due to the upper triangluarity of R). Therefore, the first component of $Q^\top\mathbf{x}$ is a multiple of the DC component. Since Q is an orthonomal transform, using all but the first row of $Q^\top$ for P_{remDC} projects out the DC component. In case of color images the same trick is applied to each channel by making P_{remDC} a block-diagonal matrix with $Q^\top$ as diagonal elements.

1.2. Rescaling the Data to Make Whitening an Volume Conserving Transform. Secondly, the data was scaled such that the whitening transform has determinant one, i.e. that the determinant of the globally scaled data is one. This is done by setting $\eta = \prod \lambda_i^{\frac{1}{2n}}$, where λ_i are the eigenvalues of the covariance matrix of the training data and n is their dimension. Therefore, the determinant of the covariance matrix of the data after scaling with $\frac{1}{\eta}$ is

$$\frac{1}{\eta^{2n}} \prod \lambda_i = \frac{\prod \lambda_i}{\left(\prod \lambda_i^{\frac{1}{2n}}\right)^{2n}} = 1.$$

Since the whitening transform consist of $D^{-\frac{1}{2}}U^\top$ with $UDU^\top = C$ (C is the determinant of the scaled data), the whitening must have determinant one due to

$$1 = \det(C) = \det(UDU^\top) = \det(D^{-\frac{1}{2}}U^\top)^2$$

Note, that the same scaling factor is used for the training and test set.

2. Measures Of Redundancy

Redundancies can be quantified by a comparison of coding costs. According to Shannon's channel coding theorem the entropy of a discrete random variable is an attainable lower bound on the coding cost for error-free encoding [1]. For the construction of such a code, it is necessary to know the true distribution of the random variable. If the assumed distribution $\hat{P}(k)$ used for the construction of an optimal code is different from the true distribution $P(k)$, the coding cost is given by the log-loss

$$\mathbb{E}_P[-\log(\hat{P}(k))] \quad = \quad -\sum_k P(k)\log\hat{P}(k) = H[k] + D_{KL}[P(k)||\hat{P}(k)]\,.$$

The Kullback-Leibler divergence quantifies the additional coding cost caused by using a model distribution different from the true one. As long as it is positive, the representation can be still compressed further, which means that there are still redundancies left.

For continuous random variables, the total amount of bits required for loss-less encoding is infinite. However, in analogy to the discrete case, we can use the Kullback-Leibler divergence of the true distribution to a given model distribution. The goal of redundancy reduction is to map a random variable Y to a new random variable $Z = f(Y)$ such that the distribution of Z is as close to a factorial distribution as possible. Thus we can use the Kullback-Leibler divergence of the true distribution to the product of its marginals to measure redundancy. This quantity is known as multi-information

$$I[\rho(\mathbf{z})] \quad = \quad D_{\mathrm{KL}}\left[\rho(\mathbf{z})||\textstyle\prod_{j=1}^{n}\rho_j(z_j)\right] \;=\; \int \rho(\mathbf{z}) \log \frac{\rho(\mathbf{z})}{\prod_{j=1}^{n}\rho_j(z_j)} d\mathbf{z}.$$

Algorithmically, redundancy can be reduced by finding a representation $Z = f(Y)$ such that a factorial model distribution $\hat{\rho}(\mathbf{z}) = \prod_{j=1}^{n}\hat{\rho}_j(z_j)$ is as close as possible to the true distribution $\rho(\mathbf{z})$. Since the multi-information $I[\rho(\mathbf{z})]$ is hard to estimate, one looks at the difference between the multi-informations of Y and $Z = f(Y)$, i.e. the quantity

$$\begin{aligned} \Delta I &= I[\rho(\mathbf{z})] - I[\varrho(\mathbf{y})] \\ &= D_{\mathrm{KL}}\left[\rho(\mathbf{z})||\textstyle\prod_{j=1}^{n}\hat{\rho}_j(z_j)\right] - D_{\mathrm{KL}}\left[\varrho(\mathbf{y})||\textstyle\prod_{j=1}^{n}\hat{\varrho}_j(y_j)\right] \;, \end{aligned}$$

where $\prod_{j=1}^{n}\hat{\varrho}_j(y_j)$ is a factorial model distribution for the representation Y. The following calculation shows that evaluating the redundancy reduction achieved with a mapping $\mathbf{z} = f(\mathbf{y})$ is equivalent to evaluating the difference between the log-loss of two particular model distributions.

Before doing the actual calculation, it is useful to define the different distributions involved and state some interrelations between them:

(1) $\rho(\mathbf{z})$ and $\varrho(\mathbf{y})$ are the true distributions of the random variables Y and $Z = f(Y)$. They are related by

$$\begin{aligned} \rho(\mathbf{z})d\mathbf{z} &= \rho(f(\mathbf{y})) \cdot \left|\det \frac{\partial z_i}{\partial y_j}\right| d\mathbf{y} = \varrho(\mathbf{y})d\mathbf{y} \\ \varrho(\mathbf{y})d\mathbf{y} &= \varrho(f^{-1}(\mathbf{z})) \cdot \left|\det \frac{\partial y_i}{\partial z_j}\right| d\mathbf{z} = \rho(\mathbf{z})d\mathbf{z} \;, \end{aligned}$$

where $\frac{\partial z_i}{\partial y_j}$ denotes the Jacobian for f and $\frac{\partial y_i}{\partial z_j}$ the Jacobian of f^{-1}. Note that $\left|\det \frac{\partial z_i}{\partial y_j}\right| = \left|\det \frac{\partial y_i}{\partial z_j}\right|^{-1}$.

(2) $\hat{\rho}(\mathbf{z}) := \prod_{j=1}^{n}\hat{\rho}_j(z_j)$, $\hat{\varrho}_f(\mathbf{y})$ and $\prod_{j=1}^{n}\hat{\varrho}_j(y_j)$ are the model distributions. $\prod_{j=1}^{n}\hat{\varrho}_j(y_j)$ is the factorial model for the representation Y. The non-factorial model distribution $\hat{\varrho}_f(\mathbf{y})$ was chosen such that the function f maps it into a factorial distribution, i.e.

$$\begin{aligned} \textstyle\prod_{j=1}^{n}\hat{\rho}_j(z_j) &\overset{\text{choice of } f}{=} \hat{\rho}(\mathbf{z}) \\ &= \hat{\rho}_f(f(\mathbf{y})) \cdot \left|\det \frac{\partial z_i}{\partial y_j}\right| \\ &= \hat{\varrho}_f(\mathbf{y}). \end{aligned}$$

Now, we can write the difference in multi-information as

$$\begin{aligned}
\Delta I &= I[\rho(\mathbf{z})] - I[\varrho(\mathbf{y})] \\
&= D_{\mathrm{KL}}\left[\rho(\mathbf{z})||\prod_{j=1}^n \hat{\rho}_j(z_j)\right] - D_{\mathrm{KL}}\left[\varrho(\mathbf{y})||\prod_{j=1}^n \hat{\varrho}_j(y_j)\right] \\
&= \mathbb{E}_\rho\left[\log \frac{\rho(\mathbf{z})}{\prod_{j=1}^n \hat{\rho}_j(z_j)}\right] - \mathbb{E}_\varrho\left[\log \frac{\varrho(\mathbf{y})}{\prod_{j=1}^n \hat{\varrho}_j(y_j)}\right] \\
&= \mathbb{E}_\varrho\left[\log \frac{\rho(f(\mathbf{y}))\cdot\left|\det \frac{\partial z_i}{\partial y_j}\right|}{\hat{\varrho}_f(\mathbf{y})}\right] - \mathbb{E}_\varrho\left[\log \frac{\varrho(\mathbf{y})}{\prod_{j=1}^n \hat{\varrho}_j(y_j)}\right] \\
&= \mathbb{E}_\varrho\left[\log \frac{\rho(f(\mathbf{y}))\cdot\left|\det \frac{\partial z_i}{\partial y_j}\right|}{\hat{\varrho}_f(\mathbf{y})} - \log \frac{\varrho(\mathbf{y})}{\prod_{j=1}^n \hat{\varrho}_j(y_j)}\right] \\
&= \mathbb{E}_\varrho\left[\log \frac{\prod_{j=1}^n \hat{\varrho}_j(y_j)}{\hat{\varrho}_f(\mathbf{y})}\cdot \frac{\overbrace{\rho(f(\mathbf{y}))\cdot\left|\det \frac{\partial z_i}{\partial y_j}\right|}^{=\varrho(\mathbf{y})}}{\varrho(\mathbf{y})}\right] \\
&= \mathbb{E}_\varrho\left[\log \frac{\prod_{j=1}^n \hat{\varrho}_j(y_j)}{\hat{\varrho}_f(\mathbf{y})}\right] \\
&= \mathbb{E}_\varrho\left[-\log \hat{\varrho}_f(\mathbf{y})\right] - \mathbb{E}_\varrho\left[-\log \prod_{j=1}^n \hat{\varrho}_j(y_j)\right].
\end{aligned}$$

Thus, if we have a model density which does not factorize with respect to $\mathbf{y}$ and we have a (possibly nonlinear) mapping $\mathbf{z} = f(\mathbf{y})$ such that the transformed model density with respect to $\mathbf{z}$ becomes factorial, we can evaluate the redundancy reduction achieved with the mapping f simply by estimating the difference in the average log-loss obtained for $\hat{\varrho}_f(\mathbf{y})$ and $\prod_{j=1}^n \hat{\varrho}_j(y_j)$.

In order to get a measure which is less dependent on the number of dimensions n we define the average log-loss (ALL) to be $\mathrm{ALL} = \frac{1}{n}\mathbb{E}[-\log \hat{\varrho}(\mathbf{y})]$ for any given model distribution $\hat{\varrho}(\mathbf{y})$.

In practice, the ALL can estimated by with the empirial mean

$$\frac{1}{n}\mathbb{E}_\varrho\left[-\log \hat{\varrho}_f(\mathbf{y})\right] \approx \frac{1}{n\cdot m}\sum_{i=1}^m -\log \hat{\varrho}_f(\mathbf{y}_i).$$

3. L_p-Spherically Symmetric Distributions

3.1. Definitions, Lemmas and Theorems. In this part, we provide the rigorous definitions, lemmas and theorems used in the paper. Most results and proofs are not new and have been collected from papers and books. Nevertheless, in many cases we adapted the original statements to our need and provided more detailed versions of the proofs. The original sources are mentioned at the respective lemmas and theorems.

Definition 1. p**-Norm**

Let $\mathbf{y} \in \mathbb{R}^n$ be an arbitrary vector. We define

$$||\mathbf{y}||_p = \left(\sum_{i=1}^{n} |y_i|^p\right)^{\frac{1}{p}}, \; p > 0$$

as the p-norm of $\mathbf{y}$. Note, that only for $p > 1$, $||\mathbf{y}||_p$ is a norm in the strict sense. However, we will also use the term "p-norm" even if only $0 < p$.

Definition 2. **p-Sphere**
The unit p-sphere $\mathbb{S}_p^{n-1}$ in n dimensions is the set of points that fulfill

$$\mathbb{S}_p^{n-1} := \{\mathbf{y} \in \mathbb{R}^n | \, ||\mathbf{y}||_p = 1, \, p > 0\}.$$

Lemma 3. ***Transformation in Radial and Spherical Coordinates*** [3]
Let $\mathbf{y} = (y_1, ...y_n)^\top$ $n \geq 2$ *be a vector in* $\mathbb{R}^n \backslash \{\mathbf{0}\}$. *Consider the transformation*

$$\mathbf{y} \mapsto (r, u_1, ..., u_{n-1}) = \left(||\mathbf{y}||_p, \frac{y_1}{||\mathbf{y}||_p}, ..., \frac{y_{n-1}}{||\mathbf{y}||_p}\right).$$

The absolute value of the determinants of the transformation on the upper and lower halfspaces

$$\begin{aligned} \mathbb{R}_+^n &:= \{\mathbf{y} \in \mathbb{R}^n | \, y_n \geq 0\} \\ \mathbb{R}_-^n &:= \{\mathbf{y} \in \mathbb{R}^n | \, y_n < 0\} \end{aligned}$$

are equal and are given by

$$|\det \mathcal{J}| = r^{n-1} \left(1 - \sum_{i=1}^{n-1} |u_i|^p\right)^{\frac{1-p}{p}}.$$

Proof. The proof is a more detailed version of the proof found in [3].
Let

$$\Delta_i := \begin{cases} 1, \, u_i \geq 0 \\ -1, \, u_i < 0. \end{cases}$$

Then we can write $|u_i| = \Delta_i u_i$. The above transformation is bijective on each of the regions $\mathbb{R}_+^n$ and $\mathbb{R}_-^n$. Let $\sigma = \text{sign}(y_n)$, then the inverse is given by

$$\begin{aligned} y_i &= u_i r \, , \, 1 \leq i \leq n-1 \\ y_n &= \sigma r \left(1 - \sum_{i=1}^{n-1} |u_i|^p\right)^{\frac{1}{p}} = \sigma r \left(1 - \sum_{i=1}^{n-1} (\Delta_i u_i)^p\right)^{\frac{1}{p}}. \end{aligned}$$

Note, that the $\sigma = \text{sign}(y_n)$ determines the halfspace in which the transformation is inverted.

First, we determine the Jacobian $\mathcal{J}$. We start with computing the derivatives

$$\begin{aligned}
\frac{\partial y_i}{\partial u_j} &= \delta_{ij} r,\ 1 \le i, j \le n-1 \\
\frac{\partial y_n}{\partial u_j} &= -\sigma r \left(1 - \sum_{i=1}^{n-1} |u_i|^p\right)^{\frac{1-p}{p}} \Delta_i^p u_i^{p-1},\ 1 \le j \le n-1 \\
\frac{\partial y_i}{\partial r} &= u_i,\ 1 \le i \le n-1 \\
\frac{\partial y_n}{\partial r} &= \sigma \left(1 - \sum_{i=1}^{n-1} (\Delta_i u_i)^p\right)^{\frac{1}{p}}.
\end{aligned}$$

Therefore, the Jacobian, is given by

$$\begin{aligned}
\mathcal{J} &= \begin{pmatrix} \frac{\partial y_1}{\partial u_1} & & \frac{\partial y_1}{\partial u_{n-1}} & \frac{\partial y_1}{\partial r} \\ \vdots & \ddots & & \vdots \\ \vdots & & \ddots & \vdots \\ \frac{\partial y_n}{\partial u_1} & & \frac{\partial y_n}{\partial u_{n-1}} & \frac{\partial y_n}{\partial r} \end{pmatrix} \\
&= \begin{pmatrix} r & 0 & \dots & u_1 \\ 0 & r & & u_2 \\ \vdots & & \ddots & \vdots \\ -\sigma r \left(1 - \sum_{i=1}^{n-1} |u_i|^p\right)^{\frac{1-p}{p}} \Delta_1^p u_1^{p-1} & \dots & \dots & \sigma \left(1 - \sum_{i=1}^{n-1} (\Delta_i u_i)^p\right)^{\frac{1}{p}} \end{pmatrix}.
\end{aligned}$$

Before actually computing the absolute value of the determinant $|\det \mathcal{J}|$, we can factor out r from the first $n-1$ columns. Furthermore, we can factor out σ from the last row. Since we take the absolute value of $\det \mathcal{J}$ and $\sigma = \{-1, 1\}$, we can remove it completely afterwards. Now we can use Laplace's formula to expand the determinant along the last column. With this, we get

$$\begin{aligned}
\frac{1}{r^{n-1}} |\det \mathcal{J}| &= \sum_{k=1}^{n-1} (-1)^{n+k} \cdot u_k \cdot (-1)^{n-1-k} \cdot -\Delta_k^p u_k^{p-1} \cdot \left(1 - \sum_{i=1}^{n-1} |u_i|^p\right)^{\frac{1-p}{p}} \\
&\quad + (-1)^{2n} \left(1 - \sum_{i=1}^{n-1} |u_i|^p\right)^{\frac{1}{p}} \\
&= \sum_{k=1}^{n-1} |u_k|^p \left(1 - \sum_{i=1}^{n-1} |u_i|^p\right)^{\frac{1-p}{p}} + \left(1 - \sum_{i=1}^{n-1} |u_i|^p\right)^{\frac{1}{p}} \\
&= \left(1 - \sum_{i=1}^{n-1} |u_i|^p\right)^{\frac{1-p}{p}} \left(\sum_{k=1}^{n-1} |u_k|^p + 1 - \sum_{k=1}^{n-1} |u_k|^p\right) \\
&= \left(1 - \sum_{i=1}^{n-1} |u_i|^p\right)^{\frac{1-p}{p}}.
\end{aligned}$$

Resolving the result for $|\det \mathcal{J}|$ completes the proof. $\square$

Theorem 4. ***p-Spherical Uniform Distribution*** [3]
Let $Y = (Y_1, ..., Y_n)^\top$ be a random vector. Let the Y_i be i.i.d. distributed with p.d.f.

$$\varrho(\mathbf{y}) \quad = \quad \frac{p^{1-\frac{1}{p}}}{2\Gamma\left(\frac{1}{p}\right)} \exp\left(-\frac{|y|^p}{p}\right), \; y \in \mathbb{R}.$$

Let $U_i = \frac{Y_i}{||Y||_p}$ for $i = 1, ..., n$. Then $\sum_{i=1}^n |U_i|^p = 1$ and the joint p.d.f of $U_1, ..., U_{n-1}$ is

$$q_u(u_1, ..., u_{n-1}) \quad = \quad \frac{p^{n-1}\Gamma\left(\frac{n}{p}\right)}{2^{n-1}\Gamma^n\left(\frac{1}{p}\right)} \left(1 - \sum_{i=1}^{n-1} |u_i|^p\right)^{\frac{1-p}{p}}$$

with $-1 < u_i < 1$, $i = 1, ..., n-1$ and $\sum_{i=1}^{n-1} |u_i|^p < 1$.

Proof. The joint p.d.f. of Y is given by

$$\varrho(\mathbf{y}) \quad = \quad \frac{p^{n-\frac{n}{p}}}{2^n\Gamma^n\left(\frac{1}{p}\right)} \exp\left(-\frac{1}{p}\sum_{i=1}^n |y_i|^p\right)$$

with $y_i \in \mathbb{R}$ and $i = 1, ..., n$. Applying the transformation

$$(y_1, ..., y_n) \quad = \quad (r, u_1, ..., u_{n-1})$$

from Lemma 3 and taking into account that each $(u_1, ..., u_{n-1})$ corresponds to $(y_1, ..., y_n)$ and $(y_1, ..., -y_n)$ we obtain

$$q(u_1, ..., u_{n-1}, r) \quad = \quad 2 \cdot \frac{p^{n-\frac{n}{p}}}{2^n\Gamma^n\left(\frac{1}{p}\right)} r^{n-1} \exp\left(-\frac{r^p}{p}\right) \left(1 - \sum_{i=1}^{n-1} |u_i|^p\right)^{\frac{1-p}{p}}.$$

By integrating out r, we obtain $q_u(u_1, ..., u_n)$:

$$\int_0^\infty q(u_1, ..., u_{n-1}, r)dr \quad = \quad \frac{p^{n-\frac{n}{p}}}{2^{n-1}\Gamma^n\left(\frac{1}{p}\right)} \left(1 - \sum_{i=1}^{n-1} |u_i|^p\right)^{\frac{1-p}{p}} \int_0^\infty r^{n-1} \exp\left(-\frac{r^p}{p}\right) dr.$$

In order to compute the integral, we use the substitution $z = \frac{r^p}{p}$ or $r = (zp)^{\frac{1}{p}}$. This yields $dr = (zp)^{\frac{1}{p}-1}dz$ and, therefore,

$$\begin{aligned}
\int_0^\infty r^{n-1} \exp\left(-\frac{r^p}{p}\right) dr \quad &= \quad \int_0^\infty (zp)^{\frac{n-1}{p}} \exp(-z)(zp)^{\frac{1-p}{p}} dz \\
&= \quad p^{\frac{n-p}{p}} \int_0^\infty z^{\frac{n}{p}-1} \exp(-z)dz \\
&= \quad p^{\frac{n-p}{p}} \Gamma\left(\frac{n}{p}\right).
\end{aligned}$$

Hence,

$$
\begin{aligned}
q_u(u_1,...,u_{n-1}) &= \int_0^\infty q(u_1,...,u_{n-1},r)dr \\
&= \frac{p^{n-\frac{n}{p}}}{2^{n-1}\Gamma^n\left(\frac{1}{p}\right)}\left(1-\sum_{i=1}^{n-1}|u_i|^p\right)^{\frac{1-p}{p}} p^{\frac{n-p}{p}}\Gamma\left(\frac{n}{p}\right) \\
&= \frac{p^{n-1}\Gamma\left(\frac{n}{p}\right)}{2^{n-1}\Gamma^n\left(\frac{1}{p}\right)}\left(1-\sum_{i=1}^{n-1}|u_i|^p\right)^{\frac{1-p}{p}}.
\end{aligned}
$$

□

In order to see, why q_u is called uniform on $\mathbb{S}_p^{n-1}$ we must observe that q_u of $\left(1-\sum_{i=1}^{n-1}|u_i|^p\right)^{\frac{1-p}{p}}$ which is due to the coordinate transformation and $\frac{p^{n-1}\Gamma\left(\frac{n}{p}\right)}{2^{n-1}\Gamma^n\left(\frac{1}{p}\right)}$ which corresponds to twice the surface area of the p-sphere (see Lemma 5). Since each $\mathbf{u}$ corresponds to two $\mathbf{y}$ before the coordinate transform (one on the upper and one on the lower halfsphere), the density of $\mathbf{u}$ in $\mathbf{y}$-coordinates corresponds to $\frac{1}{S_p^{n-1}}$ where $S_p^{n-1} = \frac{2^n\Gamma(\frac{1}{p})^n}{p^{n-1}\Gamma(\frac{n}{p})}$ is the surface area of the unit p-sphere (see Lemma 5).

As we will see in Lemma 7, $\frac{Y}{||Y||_p}$ is independent of $||Y||_p$ and, therefore, the specific form of the density ϱ does not matter as long as it is p-spherically symmetric.

Lemma 5. ***Volume and Surface of the p-Sphere***
The volume $V_p^{n-1}(r)$ of the p-Sphere with radius r is given by

$$
V_p^{n-1}(r) = \frac{r^n 2^n\Gamma(\frac{1}{p})^n}{np^{n-1}\Gamma(\frac{n}{p})}.
$$

The surface $S_p^{n-1}(r)$ is given by

$$
\begin{aligned}
S_p^{n-1}(r) &= \frac{d}{dr}V_p^{n-1}(r) \\
&= \frac{r^{n-1}2^n\Gamma(\frac{1}{p})^n}{p^{n-1}\Gamma(\frac{n}{p})}.
\end{aligned}
$$

As a convention, we leave out the argument of $V_p^{n-1}(r)$ and $S_p^{n-1}(r)$ when denoting the volume or the surface of the unit p-sphere, i.e.

$$
\begin{aligned}
V_p^{n-1} &:= V_p^{n-1}(1) \\
S_p^{n-1} &:= S_p^{n-1}(1).
\end{aligned}
$$

Proof. In order to compute the volume of the p-sphere in n-dimension, we must solve the integral $\int_{\mathbb{S}_p^{n-1}} d\mathbf{u}$. Using the volume element transformation from lemma

3, we can transform the integral into

$$\begin{aligned}
\int_{\mathbb{S}_p^{n-1}} d\mathbf{u} &= 2\int_0^r \int r^{n-1} \left(1-\sum_{i=1}^{n-1}|u_i|^p\right)^{\frac{1-p}{p}} dr\, d\mathbf{u} \\
&= 2\int_0^r r^{n-1} dr \cdot \int \left(1-\sum_{i=1}^{n-1}|u_i|^p\right)^{\frac{1-p}{p}} d\mathbf{u} \\
&= \frac{1}{n} r^n \cdot 2 \int \left(1-\sum_{i=1}^{n-1}|u_i|^p\right)^{\frac{1-p}{p}} d\mathbf{u}.
\end{aligned}$$

In theorem 4 we prove that $q(u_1,...,u_{n-1}) = \frac{p^{n-1}\Gamma\left(\frac{n}{p}\right)}{2^{n-1}\Gamma^n\left(\frac{1}{p}\right)} \left(1-\sum_{i=1}^{n-1}|u_i|^p\right)^{\frac{1-p}{p}}$ is a probability density. In particular, this means that

$$\begin{aligned}
\int q(u_1,...,u_{n-1}) d\mathbf{u} &= \frac{p^{n-1}\Gamma\left(\frac{n}{p}\right)}{2^{n-1}\Gamma^n\left(\frac{1}{p}\right)} \int \left(1-\sum_{i=1}^{n-1}|u_i|^p\right)^{\frac{1-p}{p}} d\mathbf{u} \\
&= 1
\end{aligned}$$

which is equivalent to

$$\int \left(1-\sum_{i=1}^{n-1}|u_i|^p\right)^{\frac{1-p}{p}} d\mathbf{u} = \frac{2^{n-1}\Gamma^n\left(\frac{1}{p}\right)}{p^{n-1}\Gamma\left(\frac{n}{p}\right)}.$$

Therefore,

$$\begin{aligned}
V_p^{n-1}(r) &= \int_{\mathbb{S}_p^{n-1}} d\mathbf{u} \\
&= \frac{2}{n} r^n \cdot \int \left(1-\sum_{i=1}^{n-1}|u_i|^p\right)^{\frac{1-p}{p}} d\mathbf{u} \\
&= \frac{r^n 2^n \Gamma^n\left(\frac{1}{p}\right)}{np^{n-1}\Gamma\left(\frac{n}{p}\right)}
\end{aligned}$$

Differentiation of $V_p^{n-1}(r)$ with respect to r yields the result for the surface area. $\square$

Definition 6. **L_p-Spherically Symmetric Distribution** [2] A random vector $Y = (Y_1,...,Y_n)^\top$ is said to have a L_p-spherically symmetric distribution if Y can be written as a product of two independent random variables $Y = R \cdot U$, where R is a non-negative univariate random variable with density $q_r : \mathbb{R}^+ \to \mathbb{R}^+$ and U is uniformly distributed on the unit p-sphere, i.e.

$$q_u(u_1,...,u_n) = \frac{p^{n-1}\Gamma\left(\frac{n}{p}\right)}{2^{n-1}\Gamma^n\left(\frac{1}{p}\right)} \left(1-\sum_{i=1}^{n-1}|u_i|^p\right)^{\frac{1-p}{p}}$$

(see Theorem 4).

Lemma 7. ***Probability Density Functions*** [2]
Let $Y = (Y_1, ..., Y_n)^\top$ be an n-dimensional random variable with $P\{Y = \mathbf{0}\} = 0$ and a density of the form $Y \sim \tilde{\varrho}(||\mathbf{y}||_p^p)$. Then the following three statements hold:

(1) The random variables $R = ||Y||_p$ and $U = \frac{Y}{||Y||_p}$ are independent.
(2) $U = \frac{Y}{||Y||_p}$ is uniformly distributed on the unit p-sphere $\mathbb{S}_p^{n-1}$.
(3) $R = ||Y||_p$ has a density q_r, where q_r relates to $\tilde{\varrho}$ via

$$\begin{aligned} q_r(r) &= \frac{r^{n-1}2^n\Gamma(\frac{1}{p})^n}{p^{n-1}\Gamma(\frac{n}{p})}\tilde{\varrho}(r^p) \\ &= S_p^{n-1}(r)\tilde{\varrho}(r^p),\ r > 0. \end{aligned}$$

Proof. The proof is a more detailed version of the proof found in [2].

First we transform the density of Y with the transformation of lemma 3 and obtain the new density in spherical and radial coordinates

$$\begin{aligned} q(u_1, ..., u_{n-1}, r) &= 2\left(1 - \sum_{i=1}^{n-1}|u_i|^p\right)^{\frac{1-p}{p}} \tilde{\varrho}(r^p)r^{n-1} \\ & \qquad -1 < u_i < 1,\ 1 \le i \le n-1,\ \sum_{i=1}^{n}|u_i|^p < 1. \end{aligned}$$

Since q can be written as a product of a function of r and a function of $\mathbf{u} = (u_1, ..., u_{n-1})$, U and R are independent. Thus, $||Y||_p = R$ and $U = \frac{Y}{||Y||_p}$ are independent as well.

In order to get $q_u(u_1, ..., u_{n-1})$, we must integrate out r. However, we do not know the exact form of $\tilde{\varrho}$. But since q is a probability density, we know that

$$\int_0^\infty \int q(u_1, ..., u_{n-1}, r)d\mathbf{u}dr = 1.$$

Since Y and R are independent, we can write this integral as

$$\int_0^\infty \int q(u_1, ..., u_{n-1}, r)d\mathbf{u}dr = 2\int\left(1 - \sum_{i=1}^{n-1}|u_i|^p\right)^{\frac{1-p}{p}} d\mathbf{u} \cdot \int_0^\infty \tilde{\varrho}(r^p)r^{n-1}dr.$$

From that, we can immediately derive

$$\int_0^\infty \tilde{\varrho}(r^p)r^{n-1}dr = \left(2\int\left(1 - \sum_{i=1}^{n-1}|u_i|^p\right)^{\frac{1-p}{p}} d\mathbf{u}\right)^{-1}.$$

In order to solve $\left(2\int\left(1 - \sum_{i=1}^{n-1}|u_i|^p\right)^{\frac{1-p}{p}} d\mathbf{u}\right)^{-1}$ we can use theorem 4. In this theorem, we showed that $q_u(u_1, ..., u_{n-1}) = \frac{p^{n-1}\Gamma(\frac{n}{p})}{2^{n-1}\Gamma^n(\frac{1}{p})}\left(1 - \sum_{i=1}^{n-1}|u_i|^p\right)^{\frac{1-p}{p}}$ is the uniform distribution on the p-unit sphere. In particular, we know that $\int q(u_1, ..., u_{n-1})d\mathbf{u} = 1$ and, therefore,

$$\int \left(1 - \sum_{i=1}^{n-1}|u_i|^p\right)^{\frac{1-p}{p}} d\mathbf{u} = \frac{2^{n-1}\Gamma^n\left(\frac{1}{p}\right)}{p^{n-1}\Gamma\left(\frac{n}{p}\right)}.$$

Thus,

$$\begin{aligned}
\int_0^\infty \tilde{\varrho}(r^p) r^{n-1} dr &= \left(2 \int \left(1 - \sum_{i=1}^{n-1} |u_i|^p \right)^{\frac{1-p}{p}} d\mathbf{u} \right)^{-1} \\
&= \frac{p^{n-1} \Gamma\left(\frac{n}{p}\right)}{2^n \Gamma^n \left(\frac{1}{p}\right)}
\end{aligned}$$

and

$$\begin{aligned}
q_u(u_1, ..., u_{n-1}) &= \int_0^\infty q(u_1, ..., u_{n-1}, r) dr \\
&= \left(1 - \sum_{i=1}^{n-1} |u_i|^p \right)^{\frac{1-p}{p}} \frac{p^{n-1} \Gamma\left(\frac{n}{p}\right)}{2^{n-1} \Gamma^n \left(\frac{1}{p}\right)}.
\end{aligned}$$

This shows that Y is uniformly distributed on the unit p-sphere.

The density of R can be computed by integrating out $u_1, ..., u_{n-1}$

$$\begin{aligned}
q_r(r) &= \int q(u_1, ..., u_{n-1}, r) d\mathbf{u} \\
&= \frac{2^n \Gamma^n \left(\frac{1}{p}\right)}{p^{n-1} \Gamma\left(\frac{n}{p}\right)} r^{n-1} \tilde{\varrho}(r^p),\ r > 0
\end{aligned}$$

by the same argument as in 2. This completes the proof. □

The next theorem tells us that Y is L_p-spherically symmetric distributed if and only if its density has the form $\tilde{\varrho}(||\mathbf{y}||_p^p)$.

Theorem 8. ***Form of L_p-Spherically Symmetric Distribution*** [2] *Let $Y = (Y_1, ..., Y_n)^\top$ be an n-dimensional random variable with $P\{Y = \mathbf{0}\} = 0$. Then, the density of Y has the form $\tilde{\varrho}(||\mathbf{y}||_p^p)$, where $g : \mathbb{R}_+ \to \mathbb{R}_+$ is a measurable function, if and only if $Y = RU$ is spherically symmetric distributed, with independent R and U, where R has the density*

$$q_r(r) = \frac{2^n \Gamma^n \left(\frac{1}{p}\right)}{p^{n-1} \Gamma\left(\frac{n}{p}\right)} r^{n-1} g(r^p),\ r > 0.$$

Proof. Sufficiency: Assume $Y = RU$ with independent R and U, where U is uniformly distributed on the p-sphere and R has the density q_r. Then the joint density is given by (see theorem 4):

$$\begin{aligned}
q(r, u_1, ..., u_{n-1}) &= q_r(r) \frac{p^{n-1} \Gamma\left(\frac{n}{p}\right)}{2^{n-1} \Gamma^n \left(\frac{1}{p}\right)} \left(1 - \sum_{i=1}^{n-1} |u_i|^p \right)^{\frac{1-p}{p}} \\
& \quad -1 < u_i < 1,\ 1 \leq i \leq n-1,\ \sum_{i=1}^{n-1} |u_i|^p < 1,\ r > 0.
\end{aligned}$$

Now let $y_i = ru_i$ for $1 \leq i \leq n-1$ and $|y_n| = r\left(1 - \sum_{i=1}^{n-1} |u_i|^p\right)^{\frac{1}{p}}$. We can use 3 to see that the absolute value of the determinant of the Jacobian is given by

$$\left(r^{n-1}\left(1-\sum_{i=1}^{n-1}|u_i|^p\right)^{\frac{1-p}{p}}\right)^{-1} \quad = \quad r^{1-n}\left(1-\sum_{i=1}^{n-1}|u_i|^p\right)^{\frac{p-1}{p}}.$$

Therefore,

$$\begin{aligned} p(y_1, ..., y_n) &= \frac{p^{n-1}\Gamma\left(\frac{n}{p}\right)}{2^{n-1}\Gamma^n\left(\frac{1}{p}\right)} q_r(||\mathbf{y}||_p)||\mathbf{y}||_p^{1-n} \\ &= \tilde{\varrho}(||\mathbf{y}||_p^p). \end{aligned}$$

<u>*Necessity:*</u> Assume $Y \sim \tilde{\varrho}(||Y||_p^p)$. According to lemma 7 $\frac{Y}{||Y||_p}$ and Y are independent and $\frac{Y}{||Y||_p}$ is uniformly distributed on the p-sphere. Again in lemma 7 we showed that R has the density

$$q_r(r) \quad = \quad \frac{2^n\Gamma^n\left(\frac{1}{p}\right)}{p^{n-1}\Gamma\left(\frac{n}{p}\right)} r^{n-1}\tilde{\varrho}(r^p),\ r > 0.$$

Therefore, Y is L_p-spherically symmetric distributed if and only if $Y \sim \tilde{\varrho}(||Y||_p^p)$ for some density $\tilde{\varrho}$. $\square$

3.2. **Distributions.**

3.2.1. *The p-Spherically Symmetric Distribution with Radial Mixture of Log-Normal Distribution.*

We obtain this distribution by modeling the radial component with a mixture of log-Normal distributions

$$q_r(r) \quad = \quad \sum_{k=1}^{K} \frac{\eta_k}{r\sigma_k\sqrt{2\pi}} \exp\left(-\frac{(\log r - \mu_k)^2}{2\sigma_k^2}\right).$$

Here, η_k with $\sum_k \eta_k = 1$ constitute the "prior" probability of selecting one log-Normal distribution from the mixture, and μ_k and σ_k^2 denote the mean and the variance of the kth mixture. Taking into account the uniform distribution on the p-sphere, we get

$$q(\mathbf{u}, r) \quad = \quad \left(1-\sum_{i=1}^{n-1}|u_i|^p\right)^{\frac{1-p}{p}} \frac{p^{n-1}\Gamma\left(\frac{n}{p}\right)}{2^{n-1}\Gamma^n\left(\frac{1}{p}\right)} \sum_{k=1}^{K} \frac{\eta_k}{r\sigma_k\sqrt{2\pi}} \exp\left(-\frac{(\log r - \mu_k)^2}{2\sigma_k^2}\right).$$

Reversing the coordinate transform, we obtain the distribution in Euclidean coordinates

$$\varrho(\mathbf{y}) \quad = \quad \frac{p^{n-1}\Gamma\left(\frac{n}{p}\right)}{2^{n}\Gamma^n\left(\frac{1}{p}\right)} \sum_{k=1}^{K} \frac{\eta_k}{||\mathbf{y}||_p^n\sigma_k\sqrt{2\pi}} \exp\left(-\frac{(\log ||\mathbf{y}||_p - \mu_k)^2}{2\sigma_k^2}\right).$$

Since $||\mathbf{y}||_p$ being log-Normal distributed means $\log ||\mathbf{y}||_p$ being Gaussian distributed, we can use the standard EM for a mixture of Gaussians on the log-domain to estimate the parameters of the mixture. This is justified because log (or exp) is a strictly monotonic increasing (decreasing) function and the multiplicative determinant of the Jacobian does not depend on the parameters. Therefore, the maximizing parameter values for one the mixture of log-Normal distributions also maximizes the log-likelihood of the mixture of Gaussians in the log-domain.

In order to transform the radial component into the radial component of the p-generalized distribution, we will need the cumulative distribution function, which is given by

$$\begin{aligned}
\mathcal{F}(r_0) &= \int_0^{r_0} q_r(r)dr \\
&= \int_0^{r_0} \sum_{k=1}^{K} \frac{\eta_k}{r\sigma_k\sqrt{2\pi}} \exp\left(-\frac{(\log r - \mu_k)^2}{2\sigma_k^2}\right) dr \\
&= \sum_{k=1}^{K} \eta_k \int_0^{r_0} \frac{1}{r\sigma_k\sqrt{2\pi}} \exp\left(-\frac{(\log r - \mu_k)^2}{2\sigma_k^2}\right) dr \\
&= \sum_{k=1}^{K} \eta_k \mathcal{F}_k(r_0; \mu_k, \sigma_k) \ ,
\end{aligned}$$

where $\mathcal{F}_k(r_0; \mu_k, \sigma_k)$ is simply the cumulative distribution function of the log-Normal distribution with parameters μ_k and σ_k.

3.2.2. *The p-generalized Normal distribution.* The p-generalized Normal distribution is obtained by choosing Y to be a collection of n i.i.d. random variables Y_i, each distributed according to the exponential power distribution

$$\begin{aligned}
Y_i \sim p(y) &= \frac{p}{\Gamma\left(\frac{1}{p}\right)(2\sigma^2)^{\frac{1}{p}}2} \exp\left(-\frac{|y|^p}{2\sigma^2}\right) \\
Y \sim \varrho(\mathbf{y}) = \prod_{i=1}^{n} p(y_i) &= \left(\frac{p}{\Gamma\left(\frac{1}{p}\right)(2\sigma^2)^{\frac{1}{p}}2}\right)^n \exp\left(-\frac{\sum_{i=1}^{n}|y_i|^p}{2\sigma^2}\right)
\end{aligned}$$

Since $\varrho(\mathbf{y})$ has the form $\tilde{\varrho}(||\mathbf{y}||_p^p)$, it is a proper p-spherically symmetric distribution due to Theorem 8. Note, that for the case of $p = 2$, the p-generalized Normal distribution reduces to a multivariate isotropic Gaussian. In order to compute the contrast gain control function, we need to compute the radial distribution q_r of $p(\mathbf{x})$. Transforming p according to Lemma 3 yields

$$q(r, \mathbf{u}) = \frac{p^n r^{n-1}}{\Gamma^n\left(\frac{1}{p}\right)(2\sigma)^{\frac{n}{p}}2^{n-1}} \exp\left(-\frac{r^p}{2\sigma}\right)\left(1 - \sum_{i=1}^{n-1}|u_i|^p\right)^{\frac{1-p}{p}} .$$

By integrating over $\mathbf{u}$ (see lemma 5 how to carry out the integral) we get

$$q_r(r) = \frac{p\, r^{n-1}}{\Gamma\left(\frac{n}{p}\right)(2\sigma^2)^{\frac{n}{p}}} \exp\left(-\frac{r^p}{2\sigma^2}\right)$$

In order to estimate the scale parameter σ from data $X = \{r_1, ..., r_m\} = \{||\mathbf{x}_1||_p, ..., ||\mathbf{x}_m||_p\}$, we carry out the usual procedure for maximum likelihood estimation and obtain

$$
\begin{aligned}
\frac{d}{d\sigma}\log q_r(r) &= \frac{d}{d\sigma}\left(-\frac{2n}{p}\log(\sigma)-\frac{r^p}{2\sigma^2}\right)\\
&= \frac{r^p p-2n\sigma^2}{p\sigma^3}\\
\frac{d}{d\sigma}\sum_{i=1}^{m}\log q_r(r_i) &= \sum_{i=1}^{m}\frac{r_i^p p-2n\sigma^2}{p\sigma^3}\\
&\overset{!}{=} 0.
\end{aligned}
$$

This yields

$$
\hat{\sigma} = \sqrt{\frac{p}{2mn}\sum_{i=1}^{m} r_i^p}.
$$

For the transformation of the radial component, we will also need the cumulative distribution function of

$$
q_r(r) = \frac{p\, r^{n-1}}{\Gamma\left(\frac{n}{p}\right)(2\sigma^2)^{\frac{n}{p}}}\exp\left(-\frac{r^p}{2\sigma^2}\right).
$$

It can be computed via simple integration with the substitution $y=\frac{r^p}{2\sigma^2}$

$$
\begin{aligned}
\mathcal{F}_{\mathcal{N}_p}(a) &= \int_0^a \frac{pr^{n-1}}{\Gamma\left(\frac{n}{p}\right)(2\sigma^2)^{\frac{n}{p}}}\exp\left(-\frac{r^p}{2\sigma^2}\right)dr\\
&= \frac{p}{\Gamma\left(\frac{n}{p}\right)(2\sigma^2)^{\frac{n}{p}}}\int_0^a r^{n-1}\exp\left(-\frac{r^p}{2\sigma^2}\right)dr\\
&= \frac{1}{\Gamma\left(\frac{n}{p}\right)}\int_0^{\frac{a^p}{2\sigma^2}} y^{\frac{n}{p}-1}\exp(-y)dy\\
&= \frac{\Gamma\left(\frac{n}{p},\frac{a^p}{2\sigma^2}\right)}{\Gamma\left(\frac{n}{p}\right)},
\end{aligned}
$$

where $\Gamma(z,b)=\int_0^b y^{z-1}\exp(-y)dy$ is the incomplete Γ-function.

4. Log-Likelihood of Filters under the Log-Normal Mixture Model

The log-likelihood of a basis $\mathbf{W}$ in whitened space, given a set of whitened images $\mathbf{X}=\{\mathbf{x}_1,...,\mathbf{x}_m\}$, is given by

$$
\begin{aligned}
\mathcal{L}(\mathbf{W}|\eta,\mu,\sigma) &= \sum_{i=1}^{m}\log p(\mathbf{y}_i|\eta,\mu,\sigma,\mathbf{x}_i,\mathbf{W})\\
&= m(n-1)\log p+m\log\Gamma\left(\frac{n}{p}\right)-mn\log 2-mn\log\Gamma\left(\frac{1}{p}\right)+\\
&\quad \sum_{i=1}^{m}\log\left(\sum_{k=1}^{K}\frac{\eta_k}{||\mathbf{W}\mathbf{x}||_p^n\sigma_k\sqrt{2\pi}}\exp\left(-\frac{(\log||\mathbf{W}\mathbf{x}_i||_p-\mu_k)^2}{2\sigma_k^2}\right)\right).
\end{aligned}
$$

Taking the derivative with respect to the jth row $\mathbf{w}_j$ of $\mathbf{W}$ yields

$$
\begin{aligned}
&\frac{\partial}{\partial \mathbf{w}_j} \mathcal{L}(\mathbf{W}|\eta,\mu,\sigma) \\
=\;& \sum_{i=1}^{m} \frac{\partial}{\partial \mathbf{w}_j} \log \left(\underbrace{\sum_{k=1}^{K} \frac{\eta_k}{||\mathbf{W}\mathbf{x}_i||_p^n \sigma_k \sqrt{2\pi}} \exp\left(-\frac{(\log ||\mathbf{W}\mathbf{x}_i||_p - \mu_k)^2}{2\sigma_k^2} \right)}_{=:\mathcal{L}_1(\mathbf{W}|\eta,\mu,\sigma,\mathbf{x}_i)} \right) \\
=\;& \sum_{i=1}^{m} \mathcal{L}_1(\mathbf{W}|\eta,\mu,\sigma,\mathbf{x}_i)^{-1} \cdot \sum_{k=1}^{K} \frac{\eta_k}{\sigma_k\sqrt{2\pi}} \frac{\partial}{\partial \mathbf{w}_j} \left(||\mathbf{W}\mathbf{x}_i||_p^{-n} \exp\left(-\frac{(\log ||\mathbf{W}\mathbf{x}_i||_p - \mu_k)^2}{2\sigma_k^2} \right) \right) \\
=\;& \sum_{i=1}^{m} \mathcal{L}_1(\mathbf{W}|\eta,\mu,\sigma,\mathbf{x}_i)^{-1} \times \\
& \sum_{k=1}^{K} \frac{\eta_k}{\sigma_k\sqrt{2\pi}} ||\mathbf{W}\mathbf{x}_i||_p^{-(n+1)} \exp\left(-\frac{(\log ||\mathbf{W}\mathbf{x}_i||_p - \mu_k)^2}{2\sigma_k^2} \right) \left(-n - \frac{1}{\sigma_k^2}(\log ||\mathbf{W}\mathbf{x}_i||_p - \mu_k) \right) \frac{\partial}{\partial \mathbf{w}_j} ||\mathbf{W}\mathbf{x}_i||_p \\
=\;& \sum_{i=1}^{m} \mathcal{L}_1(\mathbf{W}|\eta,\mu,\sigma,\mathbf{x}_i)^{-1} ||\mathbf{W}\mathbf{x}_i||_p^{-(n+p)} \cdot \mathbf{x}_i^\top \times \\
& \sum_{k=1}^{K} \frac{\eta_k}{\sigma_k\sqrt{2\pi}} \exp\left(-\frac{(\log ||\mathbf{W}\mathbf{x}_i||_p - \mu_k)^2}{2\sigma_k^2} \right) \left(-n - \frac{1}{\sigma_k^2}(\log ||\mathbf{W}\mathbf{x}_i||_p - \mu_k) \right) \Delta_j |\mathbf{w}_j\mathbf{x}_i|^{p-1} \;,
\end{aligned}
$$

since $\frac{\partial}{\partial \mathbf{w}_j} ||\mathbf{W}\mathbf{x}_i||_p = \frac{\partial}{\partial \mathbf{w}_j} \left(\sum_{i=1}^{n} |\mathbf{w}_i\mathbf{x}|^p\right)^{\frac{1}{p}} = ||\mathbf{W}\mathbf{x}_i||_p^{1-p} \cdot \Delta_j |\mathbf{w}_j\mathbf{x}_i|^{p-1} \cdot \mathbf{x}_i^\top$ with $\Delta_{ij} := \mathrm{sgn}(\mathbf{w}_j\mathbf{x}_i)$.

Therefore, the gradient $\frac{\partial}{\partial \mathbf{W}} \mathcal{L}(\mathbf{W}|\eta,\mu,\sigma)$ can be written as an product between two matrices $\frac{\partial}{\partial \mathbf{W}} \mathcal{L}(\mathbf{W}|\eta,\mu,\sigma) = \mathbf{A} \cdot \mathbf{B}$ with

$$
\begin{aligned}
(\mathbf{A})_{ji} &= -\Delta_{ij} |\mathbf{w}_j\mathbf{x}_i|^{p-1} \sum_{k=1}^{K} \frac{\eta_k}{\sigma_k\sqrt{2\pi}} \exp\left(-\frac{(\log ||\mathbf{W}\mathbf{x}_i||_p - \mu_k)^2}{2\sigma_k^2} \right) \left(n + \frac{1}{\sigma_k^2}(\log ||\mathbf{W}\mathbf{x}_i||_p - \mu_k) \right) \\
(\mathbf{B})_{i\ell} &= \mathcal{L}_1(\mathbf{W}|\eta,\mu,\sigma,\mathbf{x}_i)^{-1} ||\mathbf{W}\mathbf{x}_i||_p^{-(n+p)} \cdot x_{i\ell} \\
&= \left(||\mathbf{W}\mathbf{x}_i||_p^{p} \sum_{k=1}^{K} \frac{\eta_k}{\sigma_k\sqrt{2\pi}} \exp\left(-\frac{(\log ||\mathbf{W}\mathbf{x}_i||_p - \mu_k)^2}{2\sigma_k^2} \right) \right)^{-1} \cdot x_{i\ell}
\end{aligned}
$$

Absolute Difference [Bits/Comp.]			Relative Difference [% wrt. cICA]	
	Color	Gray	Color	Gray
HAD - PIX	-4.0778 ± 0.0039	$-3.1275 + -0.0040$	$92.0797 + -0.0581$	$90.8566 + -0.0854$
SYM - PIX	-4.1665 ± 0.0040	$-3.1697 + -0.0037$	94.0826 ± 0.0534	$92.0834 + -0.0876$
ICA - PIX	-4.2376 ± 0.0041	$-3.2146 + -0.0037$	95.6872 ± 0.0489	$93.3870 + -0.0823$
cHAD - PIX	-4.3516 ± 0.0055	-3.4149 ± 0.0058	98.2622 ± 0.0086	99.2077 ± 0.0103
cSYM - PIX	-4.3819 ± 0.0056	-3.4242 ± 0.0058	98.9454 ± 0.0098	99.4770 ± 0.0099
cICA - PIX	-4.4286 ± 0.0057	-3.4422 ± 0.0059	100.0000 ± 0.0000	100.0000 ± 0.0000

Table 1. Difference in ALL for gray value and color images with standard devation over ten training and test set pairs. For computational efficiency the patch size has been chosen 7×7. The columns on the left display the absolute difference to the PIX representation. The columns on the right show the percentual difference with respect to the largest reduction achieved by ICA with non-factorial model.

5. ALL Scores For Color and Gray Value Images

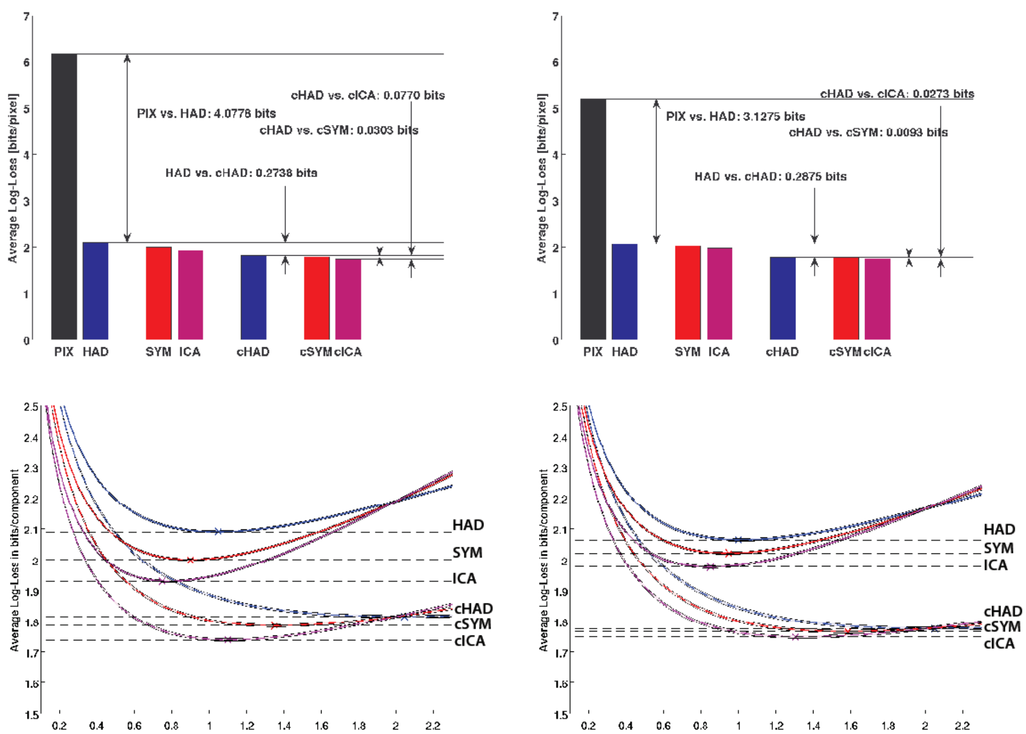

Figure 5.1. ALL in Bits per component as a function of p for achromatic (*right*) and chromatic (left) images. For computational efficiency both plots have been computed on patches of size 7×7. The slightly brighter envelope depicts the standard deviation over ten pairs of training and test sets. For further details see the respective figure in the paper.

References

[1] T.M. Cover and J.A. Thomas. *Elements of information theory*. J. Wiley & Sons, New York, 1991.

[2] A. K. Gupta and D. Song. l_p-norm spherical distribution. *Journal of Statistical Planning and Inference*, 60:241–260, 1997.

[3] D. Song and A. K. Gupta. l_p-norm uniform distribution. *Proceedings of the American Mathematical Society*, 125:595–601, 1997.