1. Direct Implementation of a Semantic Fisher Vector

We follow the derivations in Appendix A of [2] to compute the Fisher information matrix for a Dirichlet Mixture distribution. For a K mixture model with mixture weights $w_s$ and component parameters $s$, the following was shown to be a reasonable assumption.

\[
\frac{\partial p(k|x)}{\partial \alpha_s} = p(k|x) [\delta(s, k) - p(k|x)] \frac{\partial \log p(x|s)}{\partial \alpha_s} \approx 0
\]  

where $\delta(s, k)$ is an indicator function, that equals 1 if $s = k$, and is 0 otherwise. This assumption is valid for large mixture distributions, where the posterior probabilities $P(k|x)$ are very peaky, and therefore $P(k|x)P(s|x) \approx 0$ if $k \neq s$ and $P(k|x) \approx P(k|x)P(s|x)$ if $k = s$.

Second order derivative of a Dirichlet mixture log likelihood with respect to its parameters can be expressed as,

\[
\frac{\partial^2 L}{\partial \alpha_{sm} \partial \alpha_{kl}} = \frac{\partial}{\partial \alpha_{sm}} p(k|x) \left( \psi^\prime \left( \sum_l \alpha_{kl} \right) - \psi^\prime \left( \alpha_{kl} \right) + \log \pi_l \right)
\]

\[
= \left( \frac{\partial p(k|x)}{\partial \alpha_{sm}} \right) \left( \psi^\prime \left( \sum_l \alpha_{kl} \right) - \psi^\prime \left( \alpha_{kl} \right) \right)
\]

\[
+ p(k|x) \left( \frac{\partial}{\partial \alpha_{sm}} \psi^\prime \left( \sum_l \alpha_{kl} \right) - \frac{\partial}{\partial \alpha_{sm}} \psi^\prime \left( \alpha_{kl} \right) \right)
\]

\[
= 0 + p(k|x) \left( \psi^\prime \left( \sum_l \alpha_{kl} \right) - \psi^\prime \left( \alpha_{kl} \right) \delta(l, m) \right) \delta(k, s)
\]

where $\pi_l$ is the $l^{th}$ dimension of the data point $\pi$, which is a probability vector, and $\psi^\prime(x) = \frac{\partial \psi(x)}{\partial x}$ is a digamma function. The presence of $\delta(k, s)$ in the expression indicates that $\frac{\partial^2 L}{\partial \alpha_{sm} \partial \alpha_{kl}} = 0$ if $k \neq s$, that is, if the gradient is with respect to parameters of two different mixture components.

The Fisher Information matrix, therefore simplifies into the following block diagonal form.

\[
F_{lm} = E \left[ -\frac{\partial^2 \log P(\pi|\{\alpha_k, w_k\}_{k=1}^K)}{\partial \alpha_{kl} \partial \alpha_{km}} \right]
\]

\[
= E \left[ p(k|\pi) \left( \psi^\prime (\alpha_{kl}) \delta(l, m) - \psi^\prime \left( \sum_l \alpha_{kl} \right) \right) \right]
\]

\[
= w_k \left( \psi^\prime (\alpha_{kl}) \delta(l, m) - \psi^\prime \left( \sum_l \alpha_{kl} \right) \right)
\]

The matrix $F^{-1/2}$ is used to scale the DMM based Fisher scores of each image BoS (see eq (8) in the paper). The resulting image representation is referred to as a Dirichlet mixture Fisher vector (DMM FV).

References
