Scalable Face Image Retrieval Using Attribute-Enhanced Sparse Codewords

Some works only consider face part when coding the face image. However, information of skin-color, gender and hair-color are lost.

Method

Coding

Sparse Coding

D denotes the vector dictionary and v denotes the linear combination.

$$\min_{D,V} \sum_i |x^{(i)} - Dv^{(i)}|_2^2 + \lambda |v^{(i)}|_1$$

Attr-Sparse Coding

With a given attribe value, we try to split faces by their value.

Target:

- $v > 0 : (v_1, \ldots, v_n, 0, \ldots, 0)$
- $v < 0 : (0, \ldots, 0, v_{n+1}, \ldots, v_{2n})$

For multiple attribes, we concat segment from vector considering single attribe. For instance, k attribute will be denoted as

$$v = (v_1, \ldots, v_{2n \times k})$$

Hence we can modify minimize target from sparse coding:

$$\min_{D,V} \sum_i |x^{(i)} - Dv^{(i)}|_2^2 + \lambda |v^{(i)} \mathrm{Diag}(z^{(i)})|_1$$

z will be a **mask** vector where its value will be 1 or ∞ to control v to match what we want.

Indexing

Original Method

We take every face as a bag of words from dictionary.

$$c = \{d_i | v_i \neq 0\}$$

Hence the similarity can be written as:

$$S(i,j) = |c_i \cap c_j|$$

Considering Attribute

Define: $b^i_j = [f^{(i)}_a(j) > 0]$ as a bit of $b^{(i)}$

$$S(i,j) = \begin{cases} |c^{(i)} \cap c^{(j)}| & \text{if } \mathrm{ham}(b^{(i)}, b^{(j)}) \leq T \\ 0 & \text{otherwise} \end{cases}$$