Parsimonious Labeling: Supplementary Material

Puneet K. Dokania, M. Pawan Kumar
CentraleSupélec and INRIA Saclay
puneet.kumar@inria.fr, pawan.kumar@ecp.fr

1 Additional Real Data Experiments and Analysis

Recall that the energy functional of the parsimonious labeling problem is defined as:

$$E(\mathbf{x}) = \sum_{i \in \mathcal{V}} \theta_i(x_i) + \sum_{c \in \mathcal{C}} w_c \delta(\Gamma(\mathbf{x}_c)). \tag{1}$$

where $\delta()$ is the diversity function defined over the set of unique labels present in the clique $\mathbf{x}_c$. In our experiments, we frequently use the truncated linear metric. We define it below for the sake of completeness.

$$\theta_{i,j}(l_a, l_b) = \lambda \min(|l_a - l_b|, M), \forall l_a, l_b \in \mathcal{L}. \tag{2}$$

where λ is the weight associated with the metric and M is the truncation constant.

In case of real data, the high-order cliques are defined over the superpixels obtained using the mean-shift method [Comaniciu and Meer, 2002]. The clique potentials used for the experiments are the diameter diversity of the *truncated linear* metric. A truncated linear metric (equation (2)) enforces smoothness in the pairwise setting, therefore, the diameter diversity of the truncated linear metric will naturally enforce smoothness in the high-order cliques, which is a desired cue for the two applications we are dealing with.

In all the real experiments we use the following form of w_c (for the high order cliques): $w_c = \exp^{-\frac{\rho(\mathbf{x}_c)}{\sigma^2}}$, where $\rho(\mathbf{x}_c)$ is the variance of the intensities of the pixels in the clique $\mathbf{x}_c$ and σ is a hyperparameter.

In order to show the modeling capabilities of the *parsimonious labeling* we compare our results with the well known $\alpha-$expansion [Veksler, 1999], TRWS [Kolmogorov, 2006], and the Co-occ [Ladicky et al., 2010]. We also show the effect of clique sizes, which in our case are the superpixels obtained using the mean-shift algorithm, and the parameter w_c associated with the cliques, for the purpose of understanding the behaviour of the *parsimonious labeling*.

1.1 Stereo Matching

Please refer to the paper for the description of the stereo matching problem. Figures 1 and 2 shows the comparisons between different methods for the 'teddy' and 'tsukuba' examples, respectively. It can be clearly seen that the *parsimonious labeling* gives better results compared to all the other three methods. The parameter w_c can be thought of as the trade off between the influence of the pairwise and the high order cliques. Finding the best setting of w_c is very important. The effect of the parameter w_c, which is done by changing σ, is shown in the Figure 3. Similarly, the cliques have great impact on the overall result. Large cliques and high value of w_c will result in over smoothing. In order to visualize this, we show the effect of clique size in the Figure 4.

(a) Gnd Truth (b) $\alpha-$ exp (c) TRWS (d) Co-occ (e) Our Method

Figure 1: *Comparison of all the methods for the stereo matching of 'teddy'. We used the optimal setting of the parameters proposed in the well known Middlebury webpage and [Szeliski et al., 2008]. The above results are obtained using $\sigma = 10^2$ for the Co-occ and our method. Clearly, our method gives much smooth results while keeping the underlying shape intact. This is because of the cliques and the corresponding potentials (diversities) used. The diversities enforces smoothness over the cliques while σ controls this smoothness in order to avoid over smooth results.*

(a) Gnd Truth (b) $\alpha-$ exp (c) TRWS (d) Co-occ (e) Our Method

Figure 2: *Comparison of all the methods for the stereo matching of 'tsukuba'. We used the optimal setting of the parameters proposed in the well known Middlebury webpage and [Szeliski et al., 2008]. The above results are obtained using $\sigma = 10^2$ for the Co-occ and our method. We can see that the disparity obtained using our method is closest to the ground truth compared to all other methods. In our method, the background is uniform (under the table also), the camera shape is closest to the ground truth camera, and the face disparity is also closest to the ground truth compared to other methods.*

(a) $\sigma = 10^3$ (b) $\sigma = 10^4$

Figure 3: *Effect of σ in the* parsimonious labeling. *All the parameters are same except for the σ. Note that as we increase the σ, the w_c increases, which in turn results in over smoothing.*

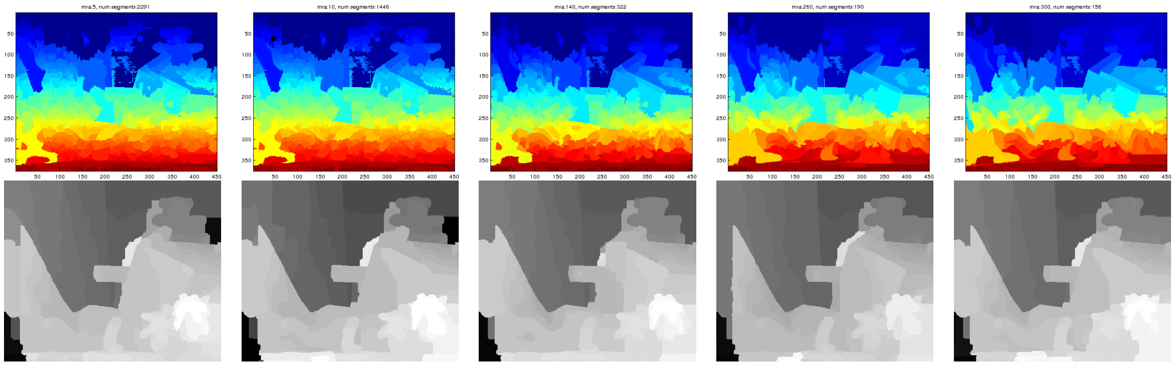


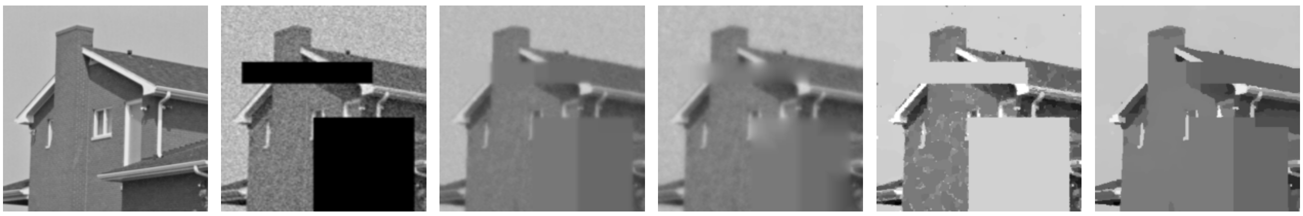
Figure 4: *Effect of clique size (superpixels). The top row shows the cliques (superpixels) used and the bottom row shows the stereo matching using these cliques. As we go from left to right, the minimum number of pixels that a superpixel must contain increases. All the other parameters are the same. In order to increase the weight* w_c*, we use high value of* σ*, which is* $\sigma = 10^5$ *in all the above cases.*

1.2 Image Inpainting and Denoising

Please refer to the paper for the description of the image inpainting and the denoising problem. Figures 5 and 6 shows the comparisons between the different methods for the 'penguin' and the 'house' examples, respectively. It can be clearly seen that the *parsimonious labeling* gives highly promising results compared to all the other methods.

(a) Original (b) Input (c) $\alpha-$ exp (d) TRWS (e) Co-occ (f) Our

Figure 5: *Comparison of all the methods for the image inpainting and denoising problem of the 'penguin'. Notice that our method recovers the hand of the penguin very smoothly. In other methods, except Co-oc, the ground is over-smooth while our method recovers the ground quite well compared to others.*



(a) Original (b) Input (c) $\alpha-$ exp (d) TRWS (e) Co-occ (f) Our

Figure 6: *Comparison of all the methods for the image inpainting and denoising problem of the 'house'.*

2 Proof of Theorems

The labeling problem. As already defined in the paper, consider a random field defined over a set of random variables $\mathbf{x} = \{x_1, \cdots, x_N\}$ arranged in a predefined lattice $\mathcal{V} = \{1, \cdots, N\}$. Each random variable can take a value from a discrete label set $\mathcal{L} = \{l_1, \cdots, l_H\}$. The energy functional corresponding to a labeling $\mathbf{x}$ is defined as:

$$E(\mathbf{x}) = \sum_{i \in \mathcal{V}} \theta_i(x_i) + \sum_{c \in \mathcal{C}} \theta_c(\mathbf{x}_c). \tag{3}$$

where $\theta_i(x_i)$ is any arbitrary unary potential, and $\theta_c(\mathbf{x}_c)$ is a clique potential for assigning the labels $\mathbf{x}_c$ to the variables in the clique c.

Notations. $\Gamma(\mathbf{x}_c)$ denotes the set of unique labels present in the clique $\mathbf{x}_c$. $\delta(\Gamma(\mathbf{x}_c))$ and $\delta^{dia}(\Gamma(\mathbf{x}_c))$ denotes the *diversity* and the *diameter diversity* of the unique labels present in the clique $\mathbf{x}_c$, respectively. $\mathcal{M} = \max_c |\mathbf{x}_c|$ is the size of the largest maximal-clique and $|\mathcal{L}|$ is the number of labels.

2.1 Multiplicative Bound of the Hierarchical Move Making Algorithm for the Hierarchical P^n Potts Model

Theorem 1. *The move making algorithm for the hierarchical P^n Potts model, Algorithm 1, gives the multiplicative bound of $\left(\frac{r}{r-1}\right) \min(\mathcal{M}, |\mathcal{L}|)$ with respect to the global minima. Here, $\mathcal{M}$ is the size of the largest maximal-clique and $|\mathcal{L}|$ is the number of labels.*

Proof. Let $\mathbf{x}^*$ be the optimal labeling of the given hierarchical P^n Potts model based labeling problem. Note that any node p in the underlying r-HST represents a cluster (subset) of labels. For each node p in the r-HST we define following sets using $\mathbf{x}^*$:

$$\begin{aligned}
\mathcal{L}^p &= \{l_i | l_i \in \mathcal{L}, i \in p\}, \\
\mathcal{V}^p &= \{x_i : x_i^* \in \mathcal{L}^p\}, \\
\mathcal{I}^p &= \{c : \mathbf{x}_c \subseteq \mathcal{V}^p\}, \\
\mathcal{B}^p &= \{c : \mathbf{x}_c \cap \mathcal{V}^p \neq \emptyset, \mathbf{x}_c \nsubseteq \mathcal{V}^p\}, \\
\mathcal{O}^p &= \{c : \mathbf{x}_c \cap \mathcal{V}^p = \emptyset\}.
\end{aligned} \tag{4}$$

In other words, $\mathcal{L}^p$ is the set of labels in the cluster at p^{th} node, $\mathcal{V}^p$ is the set of nodes whose optimal label lies in the subtree rooted at p, $\mathcal{I}^p$ is the set of cliques such that the optimal labeling lies in the subtree rooted at p, $\mathcal{B}^p$ is the set of cliques (boundary cliques) such that $\forall \mathbf{x}_c \in \mathcal{B}^p, \exists \{x_i, x_j\} \in \mathbf{x}_c : x_i^* \in \mathcal{L}^p, x_j^* \notin \mathcal{L}^p$, and $\mathcal{O}^p$ is the set of outside cliques such that the optimal assignment for all the nodes belongs to the set $\mathcal{L} \setminus \mathcal{L}^p$. Let's define $\mathbf{x}^p$ as the labeling at node p. We prove the following lemma relating $\mathbf{x}^*$ and $\mathbf{x}^p$.

Lemma 1. *Let $\mathbf{x}^p$ be the labeling at node p, $\mathbf{x}^*$ be the optimal labeling of the given hierarchical P^n Potts model, and $\delta^{dia}(\Gamma(\mathbf{x}_c^p))$ be the diameter diversity based clique potential* *defined as* $\max_{l_i,l_j \in \mathcal{L}^p} d^t(l_i,l_j), \forall p$*, where* $d^t(.,.)$ *is the tree metric defined over the given* r-HST*, then the following bound holds true at any node* p *of the* r-HST.

$$\sum_{c\in\mathcal{I}^p} \delta^{dia}(\Gamma(\mathbf{x}_c^p)) \leq \left(\frac{r}{r-1}\right) \min(\mathcal{M}, |\mathcal{L}|) \sum_{c\in\mathcal{I}^p} \delta^{dia}(\Gamma(\mathbf{x}_c^*)). \tag{5}$$

Proof. We prove the above lemma by mathematical induction. Clearly, when p is a leaf node, $x_i = p, \forall i \in \mathcal{V}$. For a non-leaf node p, we assume that the lemma holds true for the labeling $\mathbf{x}^q$ of all its children q. Given the labeling $\mathbf{x}^p$ and $\mathbf{x}^q$, we define a new labeling $\mathbf{x}^{pq}$ such that

$$\mathbf{x}^{pq} = \begin{cases} x_i^q \ \textit{if} \ x_i^* \in \mathcal{L}^q, \\ x_i^p \ \textit{otherwise.} \end{cases} \tag{6}$$

Note that $\mathbf{x}^{pq}$ lies within one α-expansion iteration away from $\mathbf{x}^p$. Since $\mathbf{x}^p$ is the local minima, we can say that

$$\begin{aligned}
E(\mathbf{x}^p|\mathcal{I}^p) + E(\mathbf{x}^p|\mathcal{B}^p) + E(\mathbf{x}^p|\mathcal{O}^p) &\leq E(\mathbf{x}^{pq}|\mathcal{I}^{pq}) + E(\mathbf{x}^{pq}|\mathcal{B}^{pq}) + E(\mathbf{x}^{pq}|\mathcal{O}^{pq}) \\
E(\mathbf{x}^p|\mathcal{I}^p) + E(\mathbf{x}^p|\mathcal{B}^p) &\leq E(\mathbf{x}^{pq}|\mathcal{I}^{pq}) + E(\mathbf{x}^{pq}|\mathcal{B}^{pq}) && (7) \\
\sum_{c\in\mathcal{I}^p} \delta^{dia}(\Gamma(\mathbf{x}_c^p)) + \sum_{c\in\mathcal{B}^p} \delta(\Gamma(\mathbf{x}_c^p)) &\leq \sum_{c\in\mathcal{I}^{pq}} \delta^{dia}(\Gamma(\mathbf{x}_c^{pq})) + \sum_{c\in\mathcal{B}^{pq}} \delta^{dia}(\Gamma(\mathbf{x}_c^{pq})) && (8) \\
\sum_{c\in\mathcal{I}^q} \delta^{dia}(\Gamma(\mathbf{x}_c^p)) + \sum_{c\in\mathcal{B}^q} \delta^{dia}(\Gamma(\mathbf{x}_c^p)) &\leq \sum_{c\in\mathcal{I}^q} \delta^{dia}(\Gamma(\mathbf{x}_c^{pq})) + \sum_{c\in\mathcal{B}^q} \delta^{dia}(\Gamma(\mathbf{x}_c^{pq})) && (9)
\end{aligned}$$

Using the mathematical induction we can write

$$\sum_{c\in\mathcal{I}^q} \delta^{dia}(\Gamma(\mathbf{x}_c^p)) + \sum_{c\in\mathcal{B}^q} \delta^{dia}(\Gamma(\mathbf{x}_c^p)) \leq \min(\mathcal{M}, |\mathcal{L}|) \left(\frac{r}{r-1}\right) \sum_{c\in\mathcal{I}^q} \delta^{dia}(\Gamma(\mathbf{x}_c^*)) + \sum_{c\in\mathcal{B}^q} \delta^{dia}(\Gamma(\mathbf{x}_c^{pq})). \tag{10}$$

Now consider a clique $c \in \mathcal{B}^q$. Let e^p be the length of edges from node p to its children q. Since $c \in \mathcal{B}^q$, there must exist atleast two nodes x_i and x_j in $\mathbf{x}_c$ such that $x_i^* \in \mathcal{L}^q$ and $x_j^* \notin \mathcal{L}^q$, therefore, by construction of r-HST

$$\delta^{dia}(\Gamma(\mathbf{x}_c^*)) \geq 2e^p \tag{11}$$

Furthermore, by the construction of $\mathbf{x}^{pq}$, $\mathcal{L}^{pq} \subseteq \mathcal{L}^p$, therefore, in worst case (leaf nodes), we can write

$$\begin{aligned}
\delta^{dia}(\Gamma(\mathbf{x}_c^{pq})) = max_{l_i,l_j\in\mathcal{L}^{pq}} d^t(l_i,l_j) &\leq 2e^p\left(1 + \frac{1}{r} + \frac{1}{r^2} + \cdots\right) \\
&= 2e^p\left(\frac{r}{r-1}\right) \\
&\leq \delta^{dia}(\Gamma(\mathbf{x}_c^*))\left(\frac{r}{r-1}\right). && (12)
\end{aligned}$$

From inequalities (10) and (12)

$$\sum_{c\in\mathcal{I}^q}\delta^{dia}(\Gamma(\mathbf{x}_c^p))+\sum_{c\in\mathcal{B}^q}\delta^{dia}(\Gamma(\mathbf{x}_c^p))\leq \min(\mathcal{M},|\mathcal{L}|)\left(\frac{r}{r-1}\right)\sum_{c\in\mathcal{I}^q}\delta^{dia}(\Gamma(\mathbf{x}_c^*))+\left(\frac{r}{r-1}\right)\sum_{c\in\mathcal{B}^q}\delta^{dia}(\Gamma(\mathbf{x}_c^*)). \tag{13}$$

In order to get the bound over the total energy we sum over all the children q of p, denoted as $\eta(p)$. Therefore, summing the inequality (13) over $\eta(p)$ we get

$$\begin{aligned}\sum_{q\in\eta(p)}\sum_{c\in\mathcal{I}^q}\delta^{dia}(\Gamma(\mathbf{x}_c^p))+\sum_{q\in\eta(p)}\sum_{c\in\mathcal{B}^q}\delta^{dia}(\Gamma(\mathbf{x}_c^p)) &\leq \min(\mathcal{M},|\mathcal{L}|)\left(\frac{r}{r-1}\right)\sum_{q\in\eta(p)}\sum_{c\in\mathcal{I}^q}\delta^{dia}(\Gamma(\mathbf{x}_c^*)) \\ &+ \left(\frac{r}{r-1}\right)\sum_{q\in\eta(p)}\sum_{c\in\mathcal{B}^q}\delta^{dia}(\Gamma(\mathbf{x}_c^*)). \quad (14)\end{aligned}$$

The LHS of the above inequality can be written as

$$\begin{aligned}\sum_{q\in\eta(p)}\sum_{c\in\mathcal{I}^q}\delta^{dia}(\Gamma(\mathbf{x}_c^p))+\sum_{q\in\eta(p)}\sum_{c\in\mathcal{B}^q}\delta^{dia}(\Gamma(\mathbf{x}_c^p)) &\geq \sum_{c\in\cup_{q\in\eta(p)}\mathcal{I}^q}\delta^{dia}(\Gamma(\mathbf{x}_c^p))+\sum_{c\in\cup_{q\in\eta(p)}\mathcal{B}^q}\delta^{dia}(\Gamma(\mathbf{x}_c^p)) \\ &= \sum_{c\in\mathcal{I}^p}\delta^{dia}(\Gamma(\mathbf{x}_c^p)). \quad (15)\end{aligned}$$

The above inequality and equality is due to the fact that $\cap_{q\in\eta(p)}\mathcal{I}^q=\emptyset$, $\cap_{q\in\eta(p)}\mathcal{B}^q$ is not necessarily an empty set, $\delta^{dia}(\Gamma(\mathbf{x}_c))\geq 0$, and $\mathcal{I}^p=\{\cup_{q\in\eta(p)}\mathcal{I}^q\}\cup\{\cup_{q\in\eta(p)}\mathcal{B}^q\}$. Now let us have a look into the second term of the RHS of the inequality (14)

$$\begin{aligned}\sum_{q\in\eta(p)}\sum_{c\in\mathcal{B}^q}\delta^{dia}(\Gamma(\mathbf{x}_c^*)) &\leq \sum_{c\in\cup_{q\in\eta(p)}\mathcal{B}^q}\min(|\eta(p)|,|\mathbf{x}_c|)\delta^{dia}(\Gamma(\mathbf{x}_c^*)) \quad (16)\\ &\leq \min\left(\max_{p\in\eta(p)}|\eta(q)|,\max_c|\mathbf{x}_c|\right)\sum_{c\in\cup_{q\in\eta(p)}\mathcal{B}^q}\delta^{dia}(\Gamma(\mathbf{x}_c^*)) \\ &= \min(\mathcal{L},|\mathcal{M}|)\sum_{c\in\cup_{q\in\eta(p)}\mathcal{B}^q}\delta^{dia}(\Gamma(\mathbf{x}_c^*)). \quad (17)\end{aligned}$$

The inequality (16) is due to the fact that $\cup_{q\in\eta(p)}\mathcal{B}^q$ can not count a clique more than $\min(|\eta(p)|,|\mathbf{x}_c|)$ times. Therefore, using the inequality (17) in the RHS of the inequality (15) we get

$$\begin{aligned}&\min(\mathcal{M},|\mathcal{L}|)\left(\frac{r}{r-1}\right)\sum_{q\in\eta(p)}\sum_{c\in\mathcal{I}^q}\delta^{dia}(\Gamma(\mathbf{x}_c^*))+\left(\frac{r}{r-1}\right)\sum_{q\in\eta(p)}\sum_{c\in\mathcal{B}^q}\delta^{dia}(\Gamma(\mathbf{x}_c^*)) \\ \leq\ &\min(\mathcal{M},|\mathcal{L}|)\left(\frac{r}{r-1}\right)\left(\sum_{c\in\cup_{q\in\eta(p)}\mathcal{I}^q}\delta^{dia}(\Gamma(\mathbf{x}_c^*))+\sum_{c\in\cup_{q\in\eta(p)}\mathcal{B}^q}\delta^{dia}(\Gamma(\mathbf{x}_c^*))\right) \\ =\ &\min(\mathcal{M},|\mathcal{L}|)\left(\frac{r}{r-1}\right)\sum_{c\in\mathcal{I}^p}\delta^{dia}(\Gamma(\mathbf{x}_c^*)). \quad (18)\end{aligned}$$

Algorithm 1 The Move Making Algorithm for the Hierarchical P^n Potts Model.

input r-HST Metric; $w_c, \forall c \in \mathcal{C}$; and $\theta_i(x_i), \forall i \in \mathcal{V}$

$\tau = T$, the leaf nodes

repeat

  for each $p \in \mathcal{N}(\tau)$ **do**

    if $|\eta(p)| = 0$, leaf node **then**

      $x_i^p = p, \forall i \in \mathcal{V}$

    else

      Fusion Move

$$\hat{\mathbf{t}}^p = \underset{\mathbf{t}^p \in \{1, \cdots, |\eta(p)|\}^N}{\operatorname{argmin}} E(\mathbf{t}^p) \tag{20}$$

      $x_i^p = x_i^{\eta(p, \hat{t}_i^p)}$.

    end if

  end for

  $\tau \leftarrow \tau - 1$

until $\tau > 0$.

Finally, using inequalities (14), (15) and (18) we get

$$\sum_{c \in \mathcal{I}^p} \delta^{dia}(\Gamma(\mathbf{x}_c^p)) \leq \min(\mathcal{M}, |\mathcal{L}|) \left(\frac{r}{r-1} \right) \sum_{c \in \mathcal{I}^p} \delta^{dia}(\Gamma(\mathbf{x}_c^*)). \tag{19}$$

□

Applying the above lemma to the root node proves the theorem. □

2.2 Multiplicative Bound of the Algorithm 2 for the Parsimonious Labeling

Theorem 2. *The move making algorithm defined in Algorithm 2 gives the multiplicative bound of* $\left(\frac{r}{r-1}\right)(|\mathcal{L}| - 1)O(\log |\mathcal{L}|)\min(\mathcal{M}, |\mathcal{L}|)$ *for the* parsimonious labeling *problem. Here,* $\mathcal{M}$ *is the size of the largest maximal-clique and* $|\mathcal{L}|$ *is the number of labels.*

Proof. Let us say that $d(.,.)$ is the induced metric of the given diversity $(\delta, \mathcal{L})$ and δ^{dia} be it's diameter diversity. We first approximate $d(.,.)$ as a mixture of r-HST metrics $d^t(.,.)$. Using Theorem 3 we get the following relationship

$$d(.,.) \leq O(\log |\mathcal{L}|) d^t(.,.). \tag{21}$$

For a given clique $\mathbf{x}_c$, using Proposition-1, we get the following relationship

$$\delta^{dia}(\Gamma(\mathbf{x}_c)) \leq \delta(\Gamma(\mathbf{x}_c)) \leq (|\Gamma(\mathbf{x}_c)| - 1)\delta^{dia}(\Gamma(\mathbf{x}_c)). \tag{22}$$

Therefore, using equations (22) and (21), we get the following inequality

$$\begin{aligned} \delta^{dia}(\Gamma(\mathbf{x}_c)) \leq \delta(\Gamma(\mathbf{x}_c)) &\leq (|\Gamma(\mathbf{x}_c)| - 1)\delta^{dia}(\Gamma(\mathbf{x}_c)) \\ &\leq O(\log |\Gamma(\mathbf{x}_c)|)(|\Gamma(\mathbf{x}_c)| - 1)\delta_t^{dia}(\Gamma(\mathbf{x}_c)). \end{aligned} \tag{23}$$

Algorithm 2 The Move Making Algorithm for the Parsimonious Labeling Problem.

input Diversity $(\mathcal{L}, \delta)$; $w_c, \forall c \in \mathcal{C}$; $\theta_i(x_i), \forall i \in \mathcal{V}$; $\mathcal{L}$; k
1: Approximate the given diversity as the mixture of k hierarchical P^n Potts model using Algorithm 3.
2: **for** each hierarchical P^n Potts model in the mixture **do**
3: Use the hierarchical move making algorithm defined in the Algorithm 1.
4: Compute energy corresponding to the solution obtained.
5: **end for**
6: Choose the solution with the minimum energy.

Algorithm 3 Diversity to Mixture of Hierarchical P^n Potts model.

input Diversity $(\mathcal{L}, \delta)$; k
1: Compute the induced metric, $d(.)$, where $d(l_i, l_j) = \delta(\{l_i, l_j\}), \forall l_i, l_j \in \mathcal{L}$.
2: Approximate $d(.)$ into mixture of k r-HST metrics $d^t(.)$ using the algorithm proposed in [Fakcharoenphol et al., 2003].
3: **for** each r-HST metrics $d^t(.)$ **do**
4: Obtain the corresponding Hierarchical P^n Potts model by defining the diameter diversity over $d^t(.)$
5: **end for**

where, $\delta_t^{dia}(\Gamma(\mathbf{x}_c))$ is the diameter diversity defined over the tree metric $d^t(.,.)$ which is obtained using the randomized algorithm [Fakcharoenphol et al., 2003] on the induced metric $d(.,.)$.

Hence, combing the inequality (23) and the previously proved Theorem 1 proves the Theorem 2.

Notice that, in case our diversity in itself is a diameter diversity, we don't need the inequality (22), therefore, the multiplicative bound reduces to $\left(\frac{r}{r-1}\right)(\log|\mathcal{L}|)\min(\mathcal{M}, |\mathcal{L}|)$. $\square$

Theorem 3. *Given any distance metric function $d(.,.)$ defined over a set of labels $\mathcal{L}$, the randomized algorithm given in [Fakcharoenphol et al., 2003] produces a mixture of r-HST tree metrics $d^t(.,.)$ such that $d(.,.) \leq O(\log|\mathcal{L}|)d^t(.,.)$.*
Proof: Please see the reference [Fakcharoenphol et al., 2003].

Proposition 1. *Let $(\mathcal{L}, \delta)$ be a diversity with induced metric space $(\mathcal{L}, d)$, then the following inequality holds $\forall \Gamma \subseteq \mathcal{L}$.*

$$\delta^{dia}(\Gamma) \leq \delta(\Gamma) \leq (|\Gamma| - 1)\delta^{dia}(\Gamma). \tag{24}$$

Proof: Please see the reference [Bryant and Tupper, 2014].

References

[Bryant and Tupper, 2014] Bryant, D. and Tupper, P. F. (2014). Diversities and the geometry of hypergraphs. In *Discrete Mathematics and Theoretical Computer Science*.

[Comaniciu and Meer, 2002] Comaniciu, D. and Meer, P. (2002). Mean shift: A robust approach toward feature space analysis. In *PAMI*.

[Fakcharoenphol et al., 2003] Fakcharoenphol, J., Rao, S., and Talwar, K. (2003). A tight bound on approximating arbitrary metrics by tree metrics. In *STOC*.

[Kolmogorov, 2006] Kolmogorov, V. (2006). Convergent tree-reweighted message passing for energy minimization. In *PAMI*.

[Ladicky et al., 2010] Ladicky, L., Russell, C., Kohli, P., and Torr, P. (2010). Graph cut based inference with co-occurrence statistics. In *ECCV*.

[Szeliski et al., 2008] Szeliski, R., Zabih, R., Scharstein, D., Veksler, O., Kolmogorov, V., Agarwala, A., Tappen, M., and Rother, C. (2008). A comparative study of energy minimization methods for markov random fields with smoothness-based priors. In *PAMI*.

[Veksler, 1999] Veksler, O. (1999). Efficient graph-based energy minimization methods in computer vision.