

KL Divergence based Agglomerative Clustering for Automated Vitiligo Grading

Mithun Das Gupta^{1*}, Srinidhi Srinivasa², Dr. Madhukara J.³, Dr. Meryl Antony³
 IBM Research Labs, Bangalore India.¹
 Ricoh Innovations Pvt. Ltd., Bangalore, India.²
 St. John's Hospital, Bangalore, India.³

mithgupt@in.ibm.com, srinidhis@ripl.ricoh.com, madhudoc2009@yahoo.com, merylantony@hotmail.com

Abstract

In this paper we present a symmetric KL divergence based agglomerative clustering framework to segment multiple levels of depigmentation in Vitiligo images. The proposed framework starts with a simple merge cost based on symmetric KL divergence. We extend the recent body of work related to Bregman divergence based agglomerative clustering and prove that the symmetric KL divergence is an upper-bound for uni-modal Gaussian distributions. This leads to a very powerful yet elegant method for bottom-up agglomerative clustering with strong theoretical guarantees. We introduce albedo and reflectance fields as features for the distance computations. We compare against other established methods to bring out possible pros and cons of the proposed method.

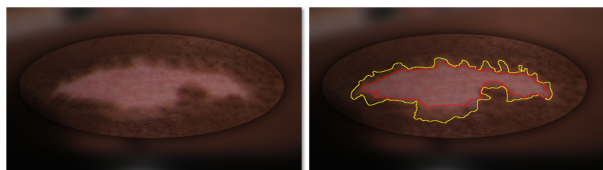


Figure 1. Vitiligo patch and its annotation by an expert. The red boundary marks the completely depigmented skin. The yellow boundary is for the partially depigmented skin. All figures are best viewed in colour.

1. Introduction

Vitiligo is the most common depigmenting disorder affecting 0.5 – 1% of the worldwide population causing disfigurement and seriously lowers quality of life. Vitiligo suffers from a lack of consensus on methods of assessment, which makes it difficult to analyse or compare the outcomes of different studies. Recently, the Vitiligo Area

Scoring Index (VASI) [22] and the Vitiligo European Task Force (VETF) [35] tools were proposed to offer more accurate measures of disease severity indexes and treatment evaluation criteria compared to simple clinical photography. VASI provides a relatively simple method to measure depigmentation over the entire patient body. Disease and treatment outcomes can be assessed using a system developed by VETF, that combines analysis of extent, stage and disease progression with respect to particular sites. Extent is evaluated by the rule of nines [20], staging is based on cutaneous and hair pigmentation, and spreading is assessed based on Wood's light examination. Objective methods to measure the spread (area) have been reported by Van Geel et al [38], but a detailed look at multiple regions of depigmentation, especially for darker skin tones (type IV and V) has largely remained an open challenge for the research community. Fig. 1 shows a typical patient with both partial as well as completely depigmented regions. Several selection criteria are of importance while considering surgical treatment in vitiligo, namely, disease type, total disease extension, resistance to non-surgical therapy, disease stability and age of the patients. Disease stability has been considered to be the most important criterion in the selection. However, no consensus exists regarding the clinical evaluation of disease activity. According to the literature, the majority of authors classified vitiligo as being stable when further progression of lesions or development of new lesions were absent in the past year. Clinical observation of lesions over time, leading to longitudinal recording of an objective score for area with respect to depigmentation seems to be an acceptable alternative. Consequently, accurate measurement of lesion area for different depigmentation is of primary importance [1]. Due to the subjectivity involved in the demarcation of the area boundary, as well as the inherent difficulty of this task, subjective measurements such as palm based eye-balling have been proposed [22]. Faint transitions between depigmented and pigmented skin patches render edge based segmentation methods obsolete. Further, illumination variations in a clinical setting can cause pixel

*This work was done when author was part of Ricoh Innovations Pvt Ltd

based segmentation algorithms to end up with many false positives. We propose a hierarchical KL divergence based cluster agglomeration approach to principally create meaningful segments to delineate patches with different levels of pigmentation, with fewer false positives.

1.1. Related work in clustering

Agglomerative (bottom-up) hierarchical clustering algorithms are important analysis tools for biological data especially gene analysis. Traditional hierarchical clustering offers three measures of distance between two clusters, namely: 1) the distance between the cluster centroids, 2) the closest points not in the same cluster and 3) the furthest points not in the same cluster. Furthermore, there are many domains [40] where clusters naturally form a hierarchy wherein clusters are themselves part of other clusters. Given pairwise dissimilarities between data points, hierarchical clustering produces a consistent result, without the need to choose initial starting positions. The dissimilarity measure is usually termed the linkage. Given the linkage, hierarchical agglomerative clustering produces a sequence of clustering assignments. At one end of the spectrum all the points are in their own clusters, and on the other end all the points are in one cluster.

A number of works, in the recent past, present agglomerative schemes for clustering with exponential families. The ideology deals mainly with the perspective of KL divergences between distributions, or the analogous goal of maximizing model likelihood, or lastly in connection to the information bottleneck method [10, 18, 25, 24, 19, 34, 9]. Additional principled smoothing techniques for divergence degeneracies were presented in [36]. Garcia et al. [19] propose a choice of one sided or symmetric Bregman divergence to cluster mixture models. Chaudhuri and McGregor [13] propose distribution clustering by KL divergence. They prove that symmetric divergences namely, Hellinger and Jensen-Shannon can be used as a relaxed metric, and the clustering obtained can be arbitrarily close to that obtained by KL divergence. The relationship between the symmetric KL divergence and the generic merge cost has not been studied yet, to the best of our knowledge. Nielsen and Nock [32] argue that the minimizer for the Jensen-Shannon divergence is the valid symmetric KL divergence which does not provide a closed form.

2. Symmetric KL Divergence based clustering

We propose symmetric Kullback-Leibler (KL) divergence (Eq. 2) between the two normally distributed clusters,

as the preferred distance metric.

$$KL_{C_i, C_j} = \frac{1}{2}(\text{tr}(\Sigma_j^{-1}\Sigma_i) + (\mu_j - \mu_i)^T \Sigma_j^{-1}(\mu_j - \mu_i) - d - \ln \frac{|\Sigma_i|}{|\Sigma_j|}) \quad (1)$$

$$D_{SKL}(C_i, C_j) = KL_{C_i, C_j} + KL_{C_j, C_i} \quad (2)$$

where C_i and C_j denote the clusters, (μ, Σ) denote the feature mean and covariance, and d is the feature dimension. The log term and the inverse covariance term are obvious bottlenecks in the formulation, which have kept researchers away from using KL divergence based cost functions. In the scenarios, where, the covariances are uniformly kept away from becoming singular, KL divergence turns out to be extremely useful divergence metric. For skin imaging with patch based processes, the covariances over the features seldom go to zero empirically. Additionally, we add a small ridge to the covariance matrix [29, 12, 21] to guarantee the boundedness of the trace as well as the log ratio terms. Optionally, as mentioned later in the text, the regions may be pre-clustered to ensure that the covariances are bounded away from being singular.

2.1. Convergence Analysis

Before presenting the convergence analysis we introduce a few notations. An exponential family [11] is a set of parametric probability distributions $\{p_F(x; \theta) | \theta \in \Theta\}$ whose probability density (or mass) can be decomposed canonically as $p_F(x; \theta) = e^{\langle t(x), \theta \rangle - F(\theta) + k(x)}$, where $t(x)$ denotes the sufficient statistics, θ the natural parameter, F the log-normalizer, and $k(x)$ the auxiliary carrier measure. $\langle x, y \rangle = x^T y$ denotes the inner product of vectors. Let $\Theta = \{\theta | \int p_F(x; \theta) dx < \infty\}$ denote the natural parameter space. It can be proved [11] that the log-normalizer $F(\theta)$ is a strictly convex and differentiable function on an open convex set Θ . The KL divergence between two members q, r of the same exponential family can be written as [31]

$$KL(q, r) = B_F(\theta_r, \theta_q) \quad (3)$$

where $B_F(\theta_r, \theta_q)$ is the Bregman divergence computed over the swapped natural parameters. Recently, Telgarsky and Dasgupta [36] have proposed agglomerative clustering with Bregman divergences. The generic merge cost for the exponential family can now be defined as

Definition 1. [36] *Let a proper convex relatively differentiable F and two finite clusters C_1, C_2 be given. Then*

$$\Delta_{F, \theta}(C_1, C_2) = \sum_{i \in \{1, 2\}} w_i B_F(\theta_{C_i}, \theta_{C_1 \cup C_2}) \quad (4)$$

where $w_i = |C_i| / (|C_1| + |C_2|)$ for $i \in [1, 2]$ and $|\cdot|$ denotes the size of the cluster.

The Bregman divergence based agglomerative clustering method iteratively selects the pair C_i, C_j which minimizes the merge cost in Eq. 4 and replaces the cluster with $C_i \cup C_j$ [36]. We claim that the symmetric KL divergence in Eq. 2 is an upper bound for the merge cost in Eq. 4 and hence minimizing the symmetric KL divergence leads to a valid clustering algorithm.

Lemma 1. *For a strictly convex function f on a closed interval $[a, b]$, let $c = w_1 a + w_2 b$ be an interior point on the interval, where $w_1 + w_2 = 1$ and $w_1, w_2 \in \mathbb{R}^+$, then*

$$(b - a)(f'(b) - f'(a)) \geq w_1 f(a) + w_2 f(b) - f(c).$$

The lemma can be proved based on the monotonic non-decreasing slope property of convex functions on a closed interval.¹

Theorem 1. $D_{SKL}(C_1, C_2) \geq \Delta_{F, \theta}(C_1, C_2)$

Proof. For notational simplicity we denote $F(\theta_{C_i})$ as F_i and θ_{C_i} as θ_i for $i = [1, 2]$ and $F(\theta_{C_1 \cup C_2})$ as F_{12} and $\theta_{C_1 \cup C_2}$ as θ_{12} . Note that $\theta_{12} = \sum_{i=[1,2]} w_i \theta_i$. Consequently,

$$\theta_{12} - \theta_1 = w_2(\theta_2 - \theta_1) \quad (5)$$

$$\theta_{12} - \theta_2 = w_1(\theta_1 - \theta_2) \quad (6)$$

$$w_1(\theta_{12} - \theta_1) + w_2(\theta_{12} - \theta_2) = 0 \quad (7)$$

$$\begin{aligned} \Delta_{F, \theta}(C_1, C_2) &= w_1 B_F(\theta_1, \theta_{12}) + w_2 B_F(\theta_2, \theta_{12}) \\ &= w_1(F_1 - F_{12} - (\theta_1 - \theta_{12})^T \nabla F_{12}) + \\ &\quad w_2(F_2 - F_{12} - (\theta_2 - \theta_{12})^T \nabla F_{12}) \\ &= (w_1 F_1 + w_2 F_2) - F_{12} \end{aligned} \quad (8)$$

where the last equality follows from $w_1 + w_2 = 1$ and Eq. 7. From Eq. 3, the weighted KL divergence can be written as

$$\begin{aligned} D_{SKL}(C_1, C_2) &= KL(C_1, C_2) + KL(C_2, C_1) \\ &= B_F(\theta_2, \theta_1) + B_F(\theta_1, \theta_2) \\ &= (F_2 - F_1 - (\theta_2 - \theta_1)^T \nabla F_1) + \\ &\quad (F_1 - F_2 - (\theta_1 - \theta_2)^T \nabla F_2) \\ &= (\theta_2 - \theta_1)^T (\nabla F_2 - \nabla F_1) \end{aligned} \quad (9)$$

where the second equality follows from Eq. 3. Combining everything, we need to show that

$$(\theta_2 - \theta_1)^T (\nabla F_2 - \nabla F_1) \geq (w_1 F_1 + w_2 F_2) - F_{12} \quad (10)$$

which follows from Lemma. 1. \square

¹Proof provided in the supplementary material.

3. Extension to Vitiligo image segmentation

We develop a Vitiligo image segmentation routine based on the bottom up hierarchical agglomerative clustering algorithm developed in the previous section. Superpixels provide a convenient primitive from which local image features can be computed. They reduce the complexity of subsequent image processing tasks and have become increasingly useful for image segmentation. Consequently, we use superpixels generated by the SLIC method proposed by Achanta et al. [8] as image primitives. Vitiligo is an epidermal (outer skin layer) disease leading to partial or total loss of coloration of skin. This leads to higher reflection from the diseased patches. The decomposition of a vitiligo patch into its albedo field intrinsically means that the patch can potentially be analysed without the high reflection constituent. We introduce albedo and shading images as features for vitiligo region segmentation according to different stages of depigmentation. Zhu and Yuille [41], in their seminal paper about region competition, had proposed albedo images for skin image segmentation. We inherit the same idea albeit with a modified albedo and shading generation algorithm to work as multiplicative features for our hierarchical clustering based segmentation algorithm.

3.1. Separation into Albedo and Shading Images

We propose a modified formulation of the method proposed by Chen and Koltun [14]. For a color image I , let A and S be the albedo and the shading (or reflectance) images respectively. Note that the shading field S is a non-linear function of surface normals and illumination [14], but for the development in this paper we restrict any further factorization. For every pixel p , we write the factorization for each channel separately as $I_p^c = A_p^c S_p^c$, where A_p^c is the albedo value and S_p^c is the shading value for the c^{th} channel. Transforming to log domain, we can write

$$\log(I_p^c) = \log(A_p^c) + \log(S_p^c) \Rightarrow i_p^c = a_p^c + s_p^c \quad (11)$$

We formulate the channelwise energy minimization problem as,

$$\begin{aligned} \min_{S^c, A^c} E(I^c) &= \overbrace{\sum_{\forall p \in I} \|(L_p + \epsilon)(i_p^c - a_p^c - s_p^c)\|^2}^{data} + \\ &\quad \lambda \overbrace{\sum_{p, q \in \tilde{\mathcal{N}}_p} \alpha_{p, q}^c \|a_p^c - a_q^c\|^2}^{reg} \end{aligned} \quad (12)$$

where λ is a relative weighting term, $L = \sum_c I^c / \sum_c$ is the luminosity (mean intensity across channels) and L_p is the luminosity at pixel p , and $\epsilon \approx 1^{-10}$ is a small constant which ensures that the data term remains well behaved for

dark pixels in the logarithmic domain. Regularization is applied only on the albedo image and the shading image is left unconstrained.

$$\alpha_{p,q}^c = \left(1 - \frac{\|ch_p^c - ch_q^c\|}{\max_{p,q \in \tilde{\mathcal{N}}_p} \|ch_p^c - ch_q^c\|} \right) \sqrt{L_p L_q} \quad (13)$$

where ch_p^c denotes the ‘rg chromaticity’ [2] of pixel p for channel c . For a multichannel image the rg chromaticity for channel c is given by $ch(I^c) = I^c / \sum_c I^c$. $\tilde{\mathcal{N}}_p$ denotes the combined local as well as the random non-local (far away) neighborhood for the pixel p . This combination preserves the local neighborhood and also looks at far away pixels in a principled manner. The number of non-local neighbors is kept slightly higher than the local neighbors.

3.2. Superpixel Generation

Achanta et al. [8] proposed the simple linear iterative clustering (SLIC) algorithm to generate superpixels from color images. Due to the simplicity and speed of this method, this has become one of the default methods to generate superpixels for color images. The CIELAB color $[l, a, b]$ and the pixel coordinate $[x, y]$ are used as the image features. A new distance metric d was introduced in [8], by simultaneously considering the image features and the size of the superpixels:

$$d = \sqrt{d_f^2 + d_{xy}^2 \frac{m^2}{S}} \quad (14)$$

where S is the sampling interval and m is the compactness of superpixels, such that larger m induces more compact superpixels. $d_f = \sqrt{\sum \nabla l^2}$ represents the distance in the color space and $d_{xy} = \sqrt{\nabla x^2}$ represents the distance in the coordinate space from the superpixel centroid. Hexagonal grids facilitate six-connectedness as opposed to four-connectedness in rectangular grids. Consequently, we adopt hexagonal units as the basic patch rather than rectangular units as proposed in [8]. The sampling interval S for clusters is modified as $\sqrt{N/((\sqrt{3}/2)K)}$ against rectangular sampling interval $\sqrt{N/K}$ where N is the number of pixels, and K is the desired number of superpixels.

Boundary recall vs super-pixel size has been a known engineering challenge in the field. For most of our experiments we keep the number of super-pixels generated fairly constant. For smaller images, this has the adverse effect of generating very small regions within the super-pixels, leading to close to singular feature covariances. As mentioned earlier, the guarantees for the KL divergence based clustering hold only if the covariances are bounded away from singular regions. As such an optional nearest neighbor based merging of the super-pixels [17], where $d = \sqrt{d_f^2 + d_{xy}^2}$ is performed to overcome the size vs boundary recall problem.

3.3. Features

The feature vector for individual pixels is a 10 dimensional vector generated by stacking the following weighted images. Let LAB be the CIELAB color image (3 channel), RGB the color image (3 channel), A the albedo image (3 channel), S the shading image (3 channel), and $L = \sum_c I^c / \sum_c I^c$ (1 channel) the luminosity of the image. The feature image I_f is generated as

$$I_f = \begin{bmatrix} LAB * (1 + \gamma S) \\ \alpha RGB * (1 + \gamma S) \\ \beta A * (1 + \gamma S) \\ \kappa L \end{bmatrix} \quad (15)$$

where α, β, γ and κ are free parameters to weight the different images and $*$ denotes channel wise multiplication. Note that these are the features over which the clustering algorithm, which is the core of this paper, works. Features mentioned in Sec. 3.2 and again in Sec. 4 are internal to the specific techniques. Based on aforementioned sections, we present an outline for the proposed method in Algorithm 1. Note that we use super-pixels and clusters interchangeably throughout this work.

Algorithm 1 Algorithm outline for vitiligo image segmentation by KL divergence based hierarchical clustering

Require RGB color image, number of final clusters c_F
 Generate albedo (A) and shading (S) images (Sec. 3.1)
 Generate super-pixels (clusters) (Sec. 3.2)
 Generate multi-dimensional feature set (Sec. 3.3)
repeat
 Generate adjacency matrix for the clusters.
 Compute pairwise affinity for neighboring super-pixels (Eq. 2) in the feature space
 Merge the 2 clusters with the lowest affinity and update cluster statistics
until number of clusters == c_F
 return final clusters

4. Physiology Guided Label Merging

While performing data annotation it was observed that the number of regions in the annotated ground truth images varied widely amongst the images collected, thereby nullifying the idea of a uniform stopping criteria for the clustering algorithms (proposed as well as comparison methods). To counter this we propose a suitable stopping criterion as $c_F = 15$ clusters for all the methods, where c_F is defined in Algorithm. 1. One of the known short comings of bottom up clustering methods is the fact that stopping criteria for the method can be extremely ad hoc. Consequently, we devised a top-down technique to merge the final few clusters to physiologically meaningful labels. All the compar-

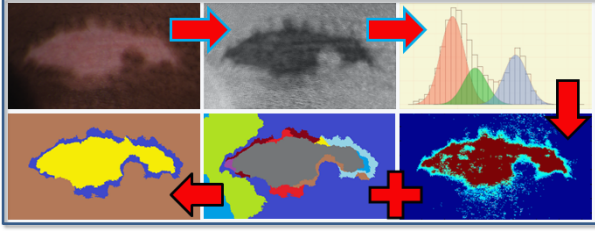


Figure 2. Label hypothesis generation and final labelling. The input image is run through the ICA engine to generate the physiological feature image. The physiological feature is used to learn a Gaussian mixture model. The model is then used to generate a label hypothesis, loosely resembling the ground-truth. The label hypothesis is used in conjunction to the segmentation output to generate the final labels.

active methods reported in this paper were stopped at the same pre-defined degree of over-segmentation ($c_F = 15$). Under-segmentation needs to be avoided strictly since it may lead to valid regions (affected) being merged into the background (normal skin). If we knew the ground truth up front, then we can label all the clusters which fall on the normal skin region as normal skin and all the clusters which fall on the affected regions as affected, for both partial and completely depigmented regions. Alternatively, we can infer a label hypothesis via physiological analysis and test our segmentation algorithm for that hypothesis.

To segregate a patch into normal and affected skin, we perform independent component analysis (ICA) based decomposition of the skin patch to the melanin and haemoglobin components [37, 30]. The primary assumption for such a decomposition is the fact that the visible color of a skin pixel is obtained by a weighted addition of melanin and haemoglobin basis colors with some additional noise. The melanin and haemoglobin bases are assumed to be independent of each other and hence the ICA principle can be applied for the decomposition. The primary decomposition can be written as

$$\mathbf{L}_p = q_p^m \mathbf{c}^m + c_p^h \mathbf{c}^h + \mathbf{c} \quad (16)$$

$$\Rightarrow [q_p^m, q_p^h] = \tilde{\mathbf{C}}^{-1} \mathbf{L}_p - \mathbf{k} \quad (17)$$

where $\mathbf{L}_p = -\log[R_p, G_p, B_p]$ is the optical density vector for the color channels at pixel p , \mathbf{c}^m and \mathbf{c}^h are pure density vectors of melanin and haemoglobin respectively, q_p^m and q_p^h are their relative quantities and \mathbf{c} and \mathbf{k} are assumed to be spatially stationary vectors accounting for other pigments and modeling errors. $\tilde{\mathbf{C}} = [\mathbf{c}^m, \mathbf{c}^h]$ as estimated by the ICA method. Linear combination of the form $f_p = \eta q_p^m + \xi q_p^h$ is used as a feature to arrive at a coarse labelling of the image. The complete data flow is shown in Fig. 2 ($[\eta, \xi] = [1, -1]$). We learn a gaussian mixture model to identify the separation bounds in the physiological feature space. The feature image is then coarsely labelled based on the likelihood to

the Gaussian mixture model [15]. This process generates a hypothesis for the possible label distribution. This hypothetical label distribution can now be used to merge the over segmented clusters obtained from the clustering algorithm. We compare each cluster with the label hypothesis and assign it to the label which has the highest overlap with it. The fused labels for all the methods with the same label hypothesis as shown in Fig. 2 bottom right panel, are shown in Fig. 3.

5. Experiments

We collected imaging data for 35 patients (15 female, 20 male) from easily accessible planar regions of the body, namely back, forearm and dorsal palm. 25 patients were unstable Vitiligo cases and the rest were clinically stable cases. The skin type distribution was, 7 with type III, 25 with type IV and 3 with type V. Planar region was preferred by the experts such that the annotations produced by them were of the highest quality. More than one site was imaged for a few patients leading to a total of 52 images. For all 52 images ground truth annotations were obtained from an expert (first expert). Additional annotations for 10 images were obtained from another (second) expert to compare inter-operator variability for the proposed method. Note that out of the 52 images only 31 had significant partially depigmented regions annotated by the first expert (Fig. 1, yellow annotation). The remaining 22 were annotated as a single completely depigmented region.

For the albedo and shading image generation part (Eq. 12), λ was set to 0.45 for the experiments reported in this paper. The superpixel generation parameters were $K = 300$ and $m = 12$. Feature computation parameters were fixed at $\{\alpha, \beta, \gamma, \kappa\} = \{0.8, 0.1, 0.4, 0.5\}$ for the experiments reported in this paper. The proposed as well as the comparison methods were implemented in C++ on an Intel 2.5GHz 64 bit CPU with 12GB RAM.

5.1. Comparative Methods

We compared the performance of our segmentation technique against three state-of-the-art bottom-up segmentation algorithms. Description of competitive methods and our efforts in extracting optimal results from them are described below. For all the methods, the algorithm stops at around 15 segments. The final segments are generated by the merging technique explained in Sec. 4.

5.1.1 Segmentation by Weighted Aggregation (SWA)

SWA[33] is a bottom-up affinity based approach for image segmentation. The algorithm is inspired by algebraic multigrid (AMG) solvers of minimization problems of heat or electric networks. It has instigated medical image segmentation technique like [15]. For a given image, a graph

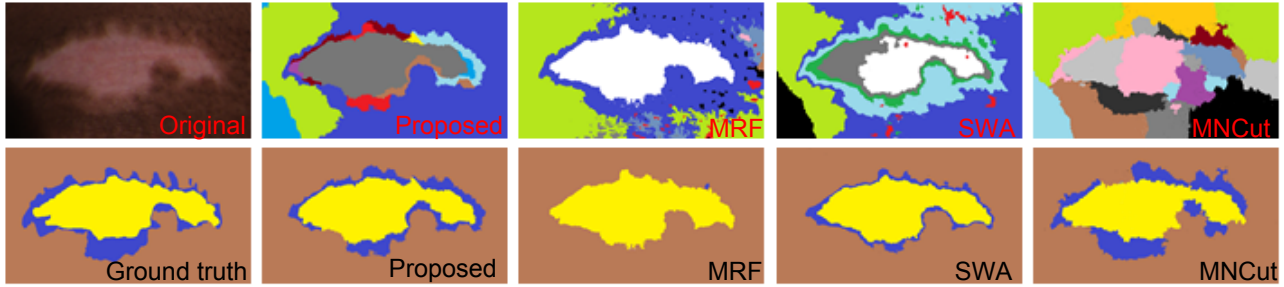


Figure 3. Evaluation paradigm. The methods over-segment the image into multiple clusters. Each segment is compared against the label hypothesis and labelled based on the maximum overlap criteria. The relabelled segments are then passed onto the numerical evaluation stage.

is constructed such that every pixel is a node in the graph and neighboring pixels are connected by an edge. A weight is associated with the edge reflecting the affinities between them. To find the minimal cuts in the graph, it is recursively coarsened using a weighted aggregation procedure in which repeatedly smaller sets of representative pixels (blocks) are selected. These representative pixels do not have to lie on a regular grid, giving rise to an irregular pyramid. The purpose of these coarsening steps is to produce smaller and smaller graphs that faithfully represent the same minimization problem. In the course of this process segments that are distinct from their environment emerge and they are detected at their appropriate size scale. After constructing the entire pyramid top down relaxation sweeps are performed to associate each pixel with appropriate segment [3].

5.1.2 Multiscale Normalized Cut (MNCut)

MNCut algorithm[16, 4] is a bottom-up spectral image segmentation technique based on multiscale graph cut principle. Typically, a fully connected graph is constructed with pixels as nodes and pairwise pixel affinities based on contour and intensity cues as edges. Pixels with strong affinity values are then clustered together into image segments, based on the normalized cut principle. However, computation of affinities for all pixel pairs is prohibitively expensive and has extensive memory requirements. A work around proposed in [16] is to compress large graphs into multiple scales, capturing image structure at increasingly larger neighborhoods to create a compact affinity matrix W with information from different scales. Generating a diagonal matrix D as $D_{i,i} = \sum_j W_{i,j}$, and having a binary state variable X for each segment l , the cost function to be maximized is : $\frac{1}{2} \sum_{l=1}^K \frac{X_l^T W X_l}{X_l^T D X_l}$, which is the optimal cut on the graph. K largest eigen vectors corresponding to the K largest eigen values are found to create K segments. Choice of the number of layers and neighborhood radii in each layer plays critical role in determination of segmenta-

tion quality and computational complexity. For our experiments, we found that two layer graphs rendered meaningful and smooth segments. $c_F = 15$ segments were generated for all images which were later refined to fewer segments using the merging technique mentioned in Sec. 4.

5.1.3 Markov Random Field Segmentation

Markov Random Fields (MRF)[28] provide a general bottom-up framework to solve classification and segmentation problems. There is widespread interest in application of MRFs for image segmentation [26, 39]. Assuming that each pixel in an image has an unknown true label representing a region, MRFs are modeled to predict the labels by striking balance between two energy terms, namely, global energy and neighborhood smoothness energy. Global energy, also known as the data-term, is characteristic of distance of a pixel from a ‘label’ in the feature space. For the experiments reported in this paper, we adopt the K-Means cost function as the data term. Local energy is a characteristic of neighborhood label similarity usually encoded by Ising’s model [5]. We used a naive K-Means clustering algorithm [23] to obtain an initial set of feature means for MRF segmentation algorithm [6]. On an average, 25 iterations were sufficient to achieve convergence.

5.2. Segmentation Quality

Empirically we found that segmenting the completely depigmented region (Fig. 1, red contour) was easier than segmenting the partially depigmented region (Fig. 1, yellow contour) as most of the methods missed the harder task completely for a few cases. Fig. 4 shows the comparison amongst all the methods tested for the easier task of segmenting the completely depigmented region. We adopt Dice’s similarity coefficient (DC) [7] to evaluate segment overlap quality, due to its wide spread acceptance within the community. Dice’s similarity coefficient is computed as

$$D = 2 \frac{|A \cap B|}{|A| + |B|} \quad (18)$$

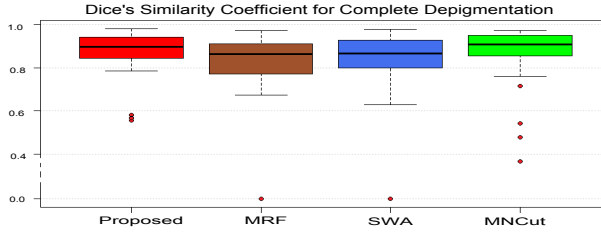


Figure 4. Boxplot for Dice’s similarity coefficient obtained for all the methods. The proposed method and MNCut surpass other methods in the easiest task of segmenting the completely depigmented region. Note that outliers for the proposed method lie higher than those for MNCut (higher is better). Also note that DC score zero means some cases were completely missed by the methods.

where A, B are two segments, $|\cdot|$ is the number of pixels and $A \cap B$ is the common region occupied by A and B . Score of 1 results from exact overlap and score 0 in case of absolutely no overlap.

We compare against the annotations obtained from the first expert for all the 52 images and report the results. Our method outperforms MRF and SWA and is comparable to MNCut with respect to segmentation accuracy. Also note that, for the easier task of segmenting the completely depigmented region, only the proposed method and MNCut were able to find a positive DC for all the images in the dataset as shown in Fig. 4. The average runtimes for the methods were MRF (1.25 s), SWA (3.92 s), MNCut (53.35 s) and proposed (8.59 s).

With the knowledge that MNCut and our proposed method are the two best performing methods for the easier task, we report the subsequent harder experiment of segmenting the partially depigmented region, with respect to these two methods only². We compare the DC scores for the partially depigmented region for the proposed method against MNCut in Fig. 5. The dotted lines represent the cases when the segmentation is completely missed by the method. Note that MNCut missed 7 cases (blue dotted lines) whereas the proposed method misses only 5 cases. Moreover, the cases missed by the proposed method were also missed by MNCut, except for one case for which MNCut registers an insignificant DC score of 0.006. We also compare the two methods with respect to the Hausdorff distance (HD) from the ground truth. Hausdorff distance is the longest of all distances from a point in one set to the closest point in the other set. Small Hausdorff distance is an indicator of strong boundary conformance while a large value represents strong boundary mismatch. It is computed as

$$d_H(A, B) = \max\left\{\sup_{a \in A} \inf_{b \in B} d(a, b), \sup_{b \in B} \inf_{a \in A} d(a, b)\right\} \quad (19)$$

²Similar experiments against all other methods are reported in the supplementary material.

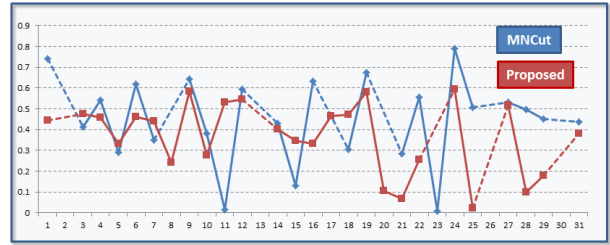


Figure 5. Dice’s similarity coefficients for the partially depigmented regions. Dotted lines represent missed cases.

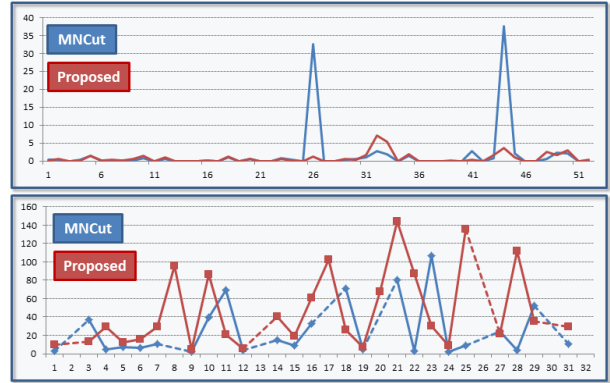


Figure 6. Hausdorff Distance comparison (lower is better). Top: complete depigmentation, bottom: partial depigmentation. Dotted lines represent missed cases.

where d is the spatial distance, a, b represent pixel locations. The comparisons against HD for the proposed method and MNCut are shown in Fig. 6. From the experiments reported, it can be safely concluded that the proposed method outperforms MNCut both in terms of run time as well as overall performance based on the number of cases missed and average accuracy. Finally, a palette showing a few challenging images with a range of skin tones and lighting variations is shown in Fig. 7.

5.3. Inter-operator Variability

The multi-level segmentation outcome, is combined with the stages recommended by the VETF([27]) to come up with an objective measurement of disease activity. We compare the disease activity score against similar scores generated by two experts. We did not notice whitening of hair in any patient, as such the inner segment (totally depigmented) was given a score 2 and outer segments with partial depigmentation were given a score 1 [27]. The combined score ($2 * \text{area}_{CD} + 1 * \text{area}_{PD}$), where CD stands for complete depigmentation and PD stands for partial depigmentation, for the two experts was compared against our method as shown in Tab. 1. The Spearman’s rank correlation for our method against expert 1 is 0.9998 and that against expert 2 is 0.9636, both with p – value < 0.05 . Note that the correlation between the two experts is also 0.9636, which

patient#	1	2	3	4	5	6	7	8	9	10
expert 1	20.827	0.293	2.818	9.614	27.362	17.034	29.552	5.200	5.493	9.307
expert 2	16.045	0.115	3.151	12.101	27.159	16.377	22.038	6.159	5.479	8.273
proposed	21.189	0.162	2.740	11.123	25.998	16.016	32.528	5.160	5.423	9.129

Table 1. Comparison of disease activity score across two experts and the proposed algorithm.

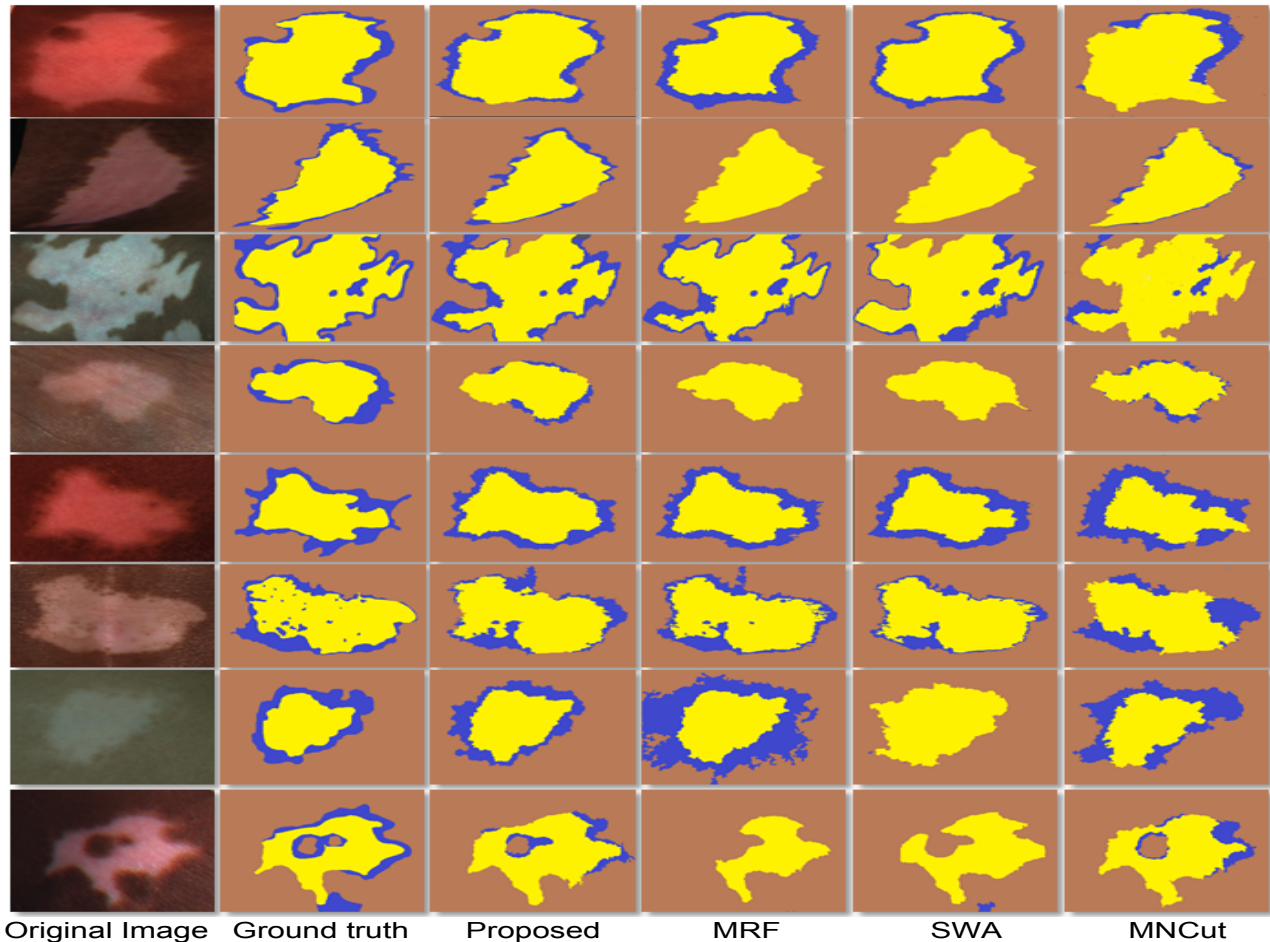


Figure 7. Final masks generated after label merging.

illustrates the reproducibility of the area based disease activity measurement technique.

6. Conclusion

In this paper, we present a novel method to segment vitiligo region with different levels of depigmentation. We show that symmetric KL divergence based clustering is a provable upper bound on Bregman divergence based clustering for unimodal non singular Gaussians. We propose a superpixel aggregation based segmentation method with augmented albedo and shading features. We further develop a physiology guided label merging technique which can be used across several agglomerative as well as graph lapla-

cian based spectral clustering methods to arrive at meaningful label configurations. We provide competitive evaluation against established methods and show that the proposed method outperforms or matches most of the existing techniques both in performance as well as run time. We statistically verify that the proposed method performs in concordance with the evaluations of two expert dermatologists on an objective disease activity score. In the future, we would like to introduce saliency based costs into the framework to identify stable regions and prune them from subsequent merging. On the broader clinical front, we would like to look into automatic identification of the VETF depigmentation stages, such that the entire pipeline can be automated with minimum user intervention.

References

- [1] http://www.espcr.org/docs/Report_VETF_Gauthier.pdf. 1
- [2] http://en.wikipedia.org/wiki/Rg_chromaticity. 4
- [3] <http://www.cse.buffalo.edu/~jcorso/r/supervoxels/>. 6
- [4] http://timotheecour.com/software/ncut_multiscale/. 6
- [5] http://en.wikipedia.org/wiki/Ising_model. 6
- [6] http://www.itk.org/Doxygen/html/group__MRFFilters.html. 6
- [7] <http://sve.bmap.ucla.edu/instructions/metrics/dice/>. 6
- [8] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Süsstrunk. SLIC Superpixels Compared to State-of-the-art Superpixel Methods. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 34(11):2274–2282, 2012. 3, 4
- [9] A. Banerjee, S. Merugu, I. S. Dhillon, and J. Ghosh. Clustering with bregman divergences. *Journal of Machine Learning Research*, 6:1705–1749, 2005. 2
- [10] C. Blundell, Y. W. Teh, and K. A. Heller. Bayesian rose trees. In P. Grnwald and P. Spirtes, editors, *UAI*, pages 65–72. AUAI Press, 2010. 2
- [11] L. D. Brown. *Fundamentals of statistical exponential families: with applications in statistical decision theory*. Institute of Mathematical Statistics, 1986. 2
- [12] O. Chapelle. Training a support vector machine in the primal. *Neural Computation*, 19:1155–1178, 2007. 2
- [13] K. Chaudhuri and A. McGregor. Finding metric structure in information theoretic clustering. In *COLT*, pages 391–402, 2008. 2
- [14] Q. Chen and V. Koltun. A simple model for intrinsic image decomposition with depth cues. In *IEEE Int. Conf. Computer Vision (ICCV)*, December 2013. 3
- [15] J. J. Corso, E. Sharon, S. Dube, S. El-Saden, U. Sinha, and A. Yuille. Efficient Multilevel Brain Tumor Segmentation with Integrated Bayesian Model Classification. *IEEE Transactions on Medical Imaging*, 27(5):629–640, 2008. 5
- [16] T. Cour, F. Benezit, and J. Shi. Spectral segmentation with multiscale graph decomposition. In *CVPR*, pages 1124–1131, 2005. 6
- [17] M. Ester, H.-P. Kriegel, J. Sander, and X. Xu. A density-based algorithm for discovering clusters in large spatial databases with noise. In *KDD*, pages 226–231. AAAI Press, 1996. 4
- [18] C. Fraley. Algorithms for model-based gaussian hierarchical clustering. *SIAM Journal on Scientific Computing*, 20:270–281, 1998. 2
- [19] V. Garcia, F. Nielsen, and R. Nock. Hierarchical gaussian mixture model. In *ICASSP*, pages 4070–4073. IEEE, 2010. 2
- [20] D. Gawkrödger, A. Ormerod, L. Shaw, I. Mauri-Sole, M. Whitton, M. Watts, A. Anstey, J. Ingham, and K. Young. Guideline for the diagnosis and management of vitiligo. *British Journal of Dermatology*, 159(5):1051–1076, 2008. 1
- [21] M. D. Gupta and J. Xiao. Non-negative matrix factorization as a feature selection tool for maximum margin classifiers. In *CVPR*, pages 2841–2848. IEEE, 2011. 2
- [22] I. Hamzavi, H. Jain, D. McLean, J. Shapiro, H. Zeng, and H. Lui. Parametric modeling of narrowband uv-b phototherapy for vitiligo using a novel quantitative tool: The vitiligo area scoring index. *Arch Dermatol*, 140(6):677–683, 2004. 1
- [23] J. A. Hartigan and M. A. Wong. Algorithm AS 136: A k-means clustering algorithm. *Applied statistics*, pages 100–108, 1979. 6
- [24] K. A. Heller and Z. Ghahramani. Bayesian hierarchical clustering. In *Proceedings of the 22Nd International Conference on Machine Learning, ICML '05*, pages 297–304, New York, NY, USA, 2005. ACM. 2
- [25] M. Iwayama and T. Tokunaga. Hierarchical bayesian clustering for automatic text classification. In *IJCAI*, pages 1322–1327. Morgan Kaufmann, 1995. 2
- [26] Z. Kato and T.-C. Pong. A markov random field image segmentation model for color textured images. *Image and Vision Computing*, 24(10):1103–1114, 2006. 6
- [27] T. Kawakami and T. Hashimoto. Disease severity indexes and treatment evaluation criteria in vitiligo. *Dermatology Research and Practice*, 2011, 2011. 7
- [28] R. Kindermann and J. L. Snell. *Markov random fields and their applications*, volume 1. American Math. Soc., Providence, RI, 1980. 6
- [29] N. D. Lawrence and G. Sanguinetti. Matching kernels through kullback-leibler divergence minimisation. Technical Report CS-04-12, The University of Sheffield, Department of Computer Science, 2004. 2
- [30] J. Lu, J. H. Manton, E. Kazmierczak, and R. Sinclair. Erythema detection in digital skin images. In *ICIP*, pages 2545–2548. IEEE, 2010. 5
- [31] F. Nielsen. Closed-form information-theoretic divergences for statistical mixtures. In *ICPR*, pages 1723–1726. IEEE, 2012. 2
- [32] F. Nielsen and R. Nock. Sided and symmetrized bregman centroids. *IEEE Transactions on Information Theory*, 55(6):2882–2904, 2009. 2
- [33] E. Sharon, M. Galun, D. Sharon, R. Basri, and A. Brandt. Hierarchy and adaptivity in segmenting visual scenes. *Nature*, 442(7104):810–813, 2006. 5
- [34] N. Slonim and N. Tishby. Agglomerative information bottleneck. In S. A. Solla, T. K. Leen, and K.-R. Müller, editors, *NIPS*, pages 617–623. The MIT Press, 1999. 2
- [35] A. Taïeb and M. Picardo. The definition and assessment of vitiligo: a consensus report of the vitiligo european task force. *Pigment Cell Research*, 20(1):27–35, 2007. 1
- [36] M. Telgarsky and S. Dasgupta. Agglomerative bregman clustering. In *ICML*. icml.cc / Omnipress, 2012. 2, 3
- [37] N. Tsumura, H. Haneishi, and Y. Miyake. Independent-component analysis of skin color image. *J. Opt. Soc. Am. A*, 16(9):2169–2176, Sep 1999. 5
- [38] N. Van Geel, Y. Vander Haeghen, K. Ongenae, and J.-M. Naeyaert. A new digital image analysis system useful for surface assessment of vitiligo lesions in transplantation studies. *Eur J Dermatol*, 14(3):150–5, 2004. 1
- [39] J. Xu, J. P. Monaco, and A. Madabhushi. Markov random field driven region-based active contour model (marcel): application to medical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2010*, pages 197–204. Springer, 2010. 6

- [40] Y. Zhao, G. Karypis, and U. Fayyad. Hierarchical clustering algorithms for document datasets. *Data Min. Knowl. Discov.*, 10:141–168, March 2005. [2](#)
- [41] S. C. Zhu and A. Yuille. Region competition: Unifying snakes, region growing, and bayes/mdl for multi-band image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 18:884–900, 1996. [3](#)