# EPITOMIC IMAGE COLORIZATION

*Yingzhen Yang*[1], *Xinqi Chu*[1], *Tian Tsong Ng*[2], *Alex Yong-Sang Chia*[2], *Jianchao Yang*[3], *Hailin Jin*[3], *Thomas S. Huang*[1]

[1]Department of Electrical and Computer Engineering, University of Illinois at Urbana-Champaign
405 North Mathews Avenue, Urbana, IL 61801, USA
[2] Institute for Infocomm Research, A*STAR, 1 Fusionopolis Way, #21-01 Connexis, Singapore 138632
[3] Adobe Research, San Jose, CA 95110, USA

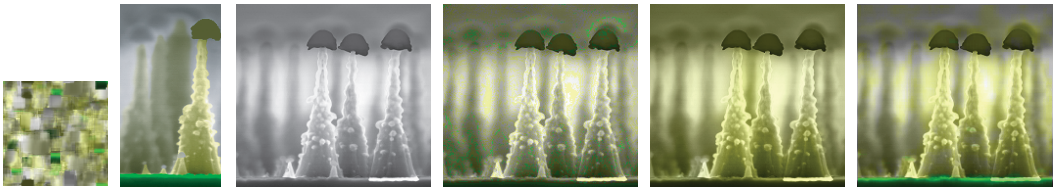

**Fig. 1:** The result of colorizing the Nano mushroom-like image acquired by the scanning electron microscopy. From left to right: the learned heterogeneous feature epitome from the reference image, the reference image, the target image, the result by Welsh et al., the result by Gupta et al., and our result.

## ABSTRACT

Image colorization adds color to grayscale images. It not only increases the visual appeal of grayscale images, but also enriches the information conveyed by scientific images that lack color information. We develop a new image colorization method, epitomic image colorization, which automatically transfers color from the reference color image to the target grayscale image by a robust feature matching scheme using a new feature representation, namely the heterogeneous feature epitome. As a generative model, heterogeneous feature epitome is a condensed representation of image appearance which is employed for measuring the dissimilarity between reference patches and target patches in a way robust to noise in the reference image. We build a Markov Random Field (MRF) model with the learned heterogeneous feature epitome from the reference image, and inference in the MRF model achieves robust feature matching for transferring color. Our method renders better colorization results than the current state-of-the-art automatic colorization methods in our experiments.

***Index Terms***— Image Colorization, Epitome, Markov Random Field

## 1. INTRODUCTION

Colorization adds color to grayscale images by assigning color values to images which only contain a grayscale channel. It not only increases the visual appeal, but also enhances the information conveyed by scientific images. For example, the grayscale images acquired by the scanning electron microscopy can be made more illustrative by adding different colors to different parts of the images. However, the manual colorization is tedious and time consuming, so it is not suitable for batch process. To overcome this problem, we propose an automatic colorization method using a new feature representation called the heterogeneous feature epitome. Figure 1 shows the colorization result for the Nano mushroom-like (the Nano-surface structure) image, where the reference image is manually colorized. The color information is transferred from the reference image to the target image automatically and faithfully by our algorithm.

Based on the amount of user intervention required, existing colorization techniques fall into two main categories: user-aided (or interactive) colorization methods [1, 2] and automatic colorization methods [3, 4]. User-aided colorization methods require users to provide color scribbles. For example, the user scribble based method in [1] requires users to draw color scribbles in the grayscale image, and the algorithm propagated the user-provided color to the whole image requiring that similar neighboring pixels should receive similar color. The method by [2] employs color scribbles for texture segmentation, and user-provided color is propagated within each segment. Such methods require extensive human labor and the quality of colorization depends on the user-provided scribbles. Using a similar color image as a reference, the automatic colorization methods perform colorization by transferring the color from the reference image to the

978-1-4799-2893-4/14/$31.00 ©2014 IEEE

grayscale image. However, the pixel-level matching based on luminance value and neighborhood statistics adopted by [3] suffers from spatial inconsistency and the user-provided swatches are required to guide the matching process in many cases. The most recent work by Gupta et al. [4] proposes a cascade feature matching scheme for matching the target superpixels to the reference superpixels. However, it is difficult for their method to produce correct feature matching in case that the same objects in the reference image and target image exhibit large change in pose or orientation. Moreover, their feature matching scheme does not handle noise or outliers in the reference image.

We propose a new automatic image colorization method, epitomic image colorization. Epitome [5] is a generative model which summarizes raw image patches into a condensed representation similar to Gaussian Mixture Models. In order to achieve feature matching robust to both noise and the large change in the pose or orientation of the objects, we propose a new patch dissimilarity measure using the heterogeneous feature epitome, where the target patches are matched to the epitome patches rather than the reference patches. This robust patch dissimilarity measure is employed to build the data likelihood term in a Markov Random Field (MRF), and the color is faithfully transferred from the reference image to the target image by inference in the MRF model, rendering a smooth feature matching. The effectiveness of our method is demonstrated by the experimental results.

## 2. FORMULATION

### 2.1. Description of Epitomic Image Colorization

Given a reference color image $\mathbf{cI}$ and the target grayscale image $\mathbf{gI}$, we learn the heterogeneous feature epitome $\mathbf{e}$ from the reference image, then perform inference in the MRF model with a robust patch dissimilarity measure by $\mathbf{e}$ so as to transfer the color information from patches of $\mathbf{cI}$ to the corresponding grayscale patches of $\mathbf{gI}$. We will illustrate the learning and inference process in detail in the following text.

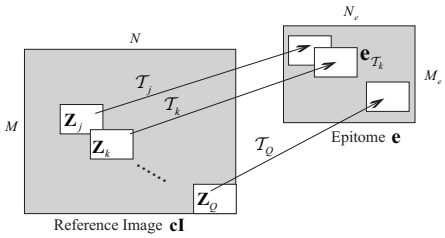


**Fig. 2:** The hidden mapping $\mathcal{T}_k$ maps the image patch $\mathbf{Z}_k$ to its corresponding epitome patch $\mathbf{e}_{\mathcal{T}_k}$ of the same size, and $\mathbf{Z}_k$ can be mapped to any possible epitome patch in the epitome $\mathbf{e}$.

### 2.2. Learning Epitome with A Single Feature

Being a generative model, the epitome of an image summarizes the raw image patches into a condensed representation of a size smaller than the original image, and it approaches this goal in a manner similar to Gaussian Mixture Models (GMM) [6]. Epitome differs from GMM in that the parameters (mean and variance) of the Gaussian components can be overlapping with each other[5], so as to improve the representation capability of finite parameter space [7, 8]. The epitome $\mathbf{e}$ of an image $I$ contains three parameters, $\mathbf{e} = (\boldsymbol{\mu}, \boldsymbol{\phi}, \boldsymbol{\pi})$, with $\boldsymbol{\mu}$ and $\boldsymbol{\phi}$ representing the Gaussian mean and variance respectively. Suppose $Q$ patches are densely sampled from the reference image, i.e. $\{\mathbf{Z}_k\}_{k=1}^{Q}$. Each patch $\mathbf{Z}_k$ is associated with a hidden mapping $\mathcal{T}_k$ which maps the image patch $\mathbf{Z}_k$ to the epitome patch $\mathbf{e}_{\mathcal{T}_k} = (\boldsymbol{\mu}_{\mathcal{T}_k}, \boldsymbol{\phi}_{\mathcal{T}_k})$ . $\boldsymbol{\pi}$ indicates the prior distribution of the hidden mapping. All the $Q$ patches are generated independently from the epitome and the corresponding hidden mappings as below:

$$\prod_{k=1}^{Q} p(\{\mathbf{Z}_k\}_{k=1}^{Q}|\{\mathcal{T}_k\}_{k=1}^{Q}, \mathbf{e}) = \prod_{k=1}^{Q} p(\mathbf{Z}_k|\mathcal{T}_k, \mathbf{e}) = \prod_{k=1}^{Q} \mathcal{N}(\mathbf{Z}_k; \boldsymbol{\mu}_{\mathcal{T}_k}, \boldsymbol{\phi}_{\mathcal{T}_k})$$

$p(\cdot)$ is the probability function, $\mathcal{N}(\cdot; \mu, \phi)$ represents a Gaussian distribution with mean $\mu$ and variance $\phi$. Based on the above formula and illustrated by Figure 2, the hidden mapping $\mathcal{T}_k$ is a hidden variable that indicates the location of the epitome patch from which the observed image patch $\mathbf{Z}_k$ is generated, and it behaves similar to the hidden variable in the Gaussian mixture models that specifies the Gaussian component from which a specific data point is generated. The epitome $\mathbf{e}$ is obtained by maximizing the log likelihood function:

$$\mathbf{e} = \arg\max_{\hat{\mathbf{e}}} \log p\left(\{\mathbf{Z}_k\}_{k=1}^{Q}|\hat{\mathbf{e}}\right), \tag{1}$$

with the Expectation-Maximization (EM) algorithm [9, 10].

### 2.3. Heterogeneous Feature Epitome

The above learning process is applicable for a single type of feature of $\mathbf{cI}$ (the pixel colors), and it can be extended to learning the epitome with heterogeneous features for a more robust feature representation. We extract three types of features from the sampled patches, i.e. the YIQ channels, the dense SIFT feature [11] and the rotation invariant Local Binary Pattern (LBP) [12]. We then learn the color epitome $\mathbf{e}^{YIQ}$, the dense SIFT epitome $\mathbf{e}^{SIFT}$ and the LBP epitome $\mathbf{e}^{LBP}$ jointly by sharing the same hidden mapping:

$$p(\mathbf{Z}_k|\mathcal{T}_k, \mathbf{e}) = p(\mathbf{Z}_k^{YIQ}|\mathcal{T}_k, \mathbf{e}^{YIQ})^{\lambda_{col}} p(\mathbf{Z}_k^{LBP}|\mathcal{T}_k, \mathbf{e}^{BBP})^{\lambda_{lbp}} \\ p(\mathbf{Z}_k^{SIFT}|\mathcal{T}_k, \mathbf{e}^{SIFT})^{1-\lambda_{col}-\lambda_{lbp}} \tag{2}$$

where $0 \leq \lambda_{col}, \lambda_{lbp} \leq 1$ are parameters balancing the preference for the color and the LBP feature, and in this way we obtain the heterogeneous feature epitome $\mathbf{e} = \left(\mathbf{e}^{YIQ}, \mathbf{e}^{SIFT}, \mathbf{e}^{LBP}\right)$ for our colorization method.

### 2.4. Robust Patch Dissimilarity Measure via Epitome

Let $\hat{Q}$ patches $\{\hat{\mathbf{Z}}_k\}_{k=1}^{\hat{Q}}$ be densely sampled from the target image $\mathbf{gI}$ (these patches cover the entire $\mathbf{gI}$). We propose the following function measuring the dissimilarity between the target patch $\hat{\mathbf{Z}}_i$ and the reference patch $\mathbf{Z}_j$ with the heterogeneous feature epitome $\mathbf{e}$ learned from $\mathbf{cI}$:

$$\mathcal{D}_{\mathbf{e}}\left(\hat{\mathbf{Z}}_i, \mathbf{Z}_j\right) = 1 - p(\hat{\mathcal{T}}_i^* | \mathbf{Z}_j, \mathbf{e}) \tag{3}$$

where $\hat{\mathcal{T}}_i^*$ is the most probable mapping for $\hat{\mathbf{Z}}_i$:

$$\hat{\mathcal{T}}_i^* = \arg\max_{\hat{\mathcal{T}}_i} p\left(\hat{\mathcal{T}}_i | \hat{\mathbf{Z}}_i, \mathbf{e}\right) \tag{4}$$

By (3), the dissimilarity between $\hat{\mathbf{Z}}_i$ and $\mathbf{Z}_j$ is inversely proportional to the posterior of the most probable mapping for $\hat{\mathbf{Z}}_i$ conditioned on the reference patch $\mathbf{Z}_j$, which improves the robustness to noise or outlier. To see this, suppose $\mathbf{Z}_j$ is an outlier or suffering from the noise in $\mathbf{cI}$, then all the posteriors $\{p(\mathcal{T}_l | \mathbf{Z}_j, \mathbf{e}, \boldsymbol{\pi})\}_{l=1}^{T}$ are small, and $\mathcal{D}_{\mathbf{e}}\left(\hat{\mathbf{Z}}_i, \mathbf{Z}_j\right)$ is large for all the target patches $\left\{\hat{\mathbf{Z}}_i\right\}$. It follows that the matching reference patch tends not to be $\mathbf{Z}_j$ for any target patch $\hat{\mathbf{Z}}_i$, revealing robustness in feature matching. Moreover, in case of large change in the pose of the objects and an associated target patch cannot have a match in $\mathbf{cI}$, it can still find a reliable match in the epitome by (4) with high probability, since each epitome patch, by its Gaussian mean and variance, summarizes a batch of similar raw patches in $\mathbf{cI}$.

### 2.5. Colorization by MRF Inference

The target image $\mathbf{gI}$ is colorized by inferring the optimal matching patches in the reference image for all the patches of the target image. In order to obtain a smooth feature matching, we build a MRF model comprising random variables $\mathbf{X} = \{\mathbf{x}_i\}_{i=1}^{\hat{Q}}$ where each $\mathbf{x}_i$ corresponds to the patch $\hat{\mathbf{Z}}_i$. The label set for $\mathbf{x}_i$ (the values $\mathbf{x}_i$ can take) is all the patches from the reference image $\mathbf{cI}$, namely $\mathcal{L}_{\mathbf{x}_i} = \{\mathbf{Z}_k\}_{k=1}^{Q}$. Note that $\mathbf{x}_i$ indicates the matching reference patch for the target patch $\hat{\mathbf{Z}}_i$. The energy function defined on the MRF model admits the following widely-used form comprising the data likelihood term (or the unary term $E_{data}$) and the pairwise term $E_{smooth}$:

$$E(\mathbf{X}) = E_{data}(\mathbf{X}) + E_{smooth}(\mathbf{X}) \tag{5}$$

$$E_{data}(\mathbf{X}) = \sum_{i=1}^{\hat{Q}} D_i(\mathbf{x}_i) \quad E_{smooth}(\mathbf{X}) = \sum_{(i,j)\in\mathcal{N}} V(\mathbf{x}_i, \mathbf{x}_j)$$

where $\mathcal{N}$ is the set of neighboring variables. $D_i(\mathbf{x}_i = \mathbf{Z}_j)$ in the unary term measures the dissimilarity between the target patch $\hat{\mathbf{Z}}_i$ and each reference patch $\mathbf{Z}_j$ for $1 \le j \le Q$, and we adopt the robust patch dissimilarity measure (3), i.e. $D_i(\mathbf{Z}_j) = \mathcal{D}_{\mathbf{e}}\left(\hat{\mathbf{Z}}_i, \mathbf{Z}_j\right)$. The pairwise term encourages neighboring nodes to take similar labels, resulting in a smooth labeling. $V(\mathbf{x}_i, \mathbf{x}_j)$ represents the dissimilarity between two nodes $\mathbf{x}_i$ and $\mathbf{x}_j$, which is defined below:

$$V(\mathbf{x}_i, \mathbf{x}_j) = C_{col} \cdot \alpha_{col} \left\|\mathbf{x}_i^{YIQ} - \mathbf{x}_j^{YIQ}\right\| + C_{pos} \cdot \alpha_{pos} \left\|\mathbf{x}_i^{pos} - \mathbf{x}_j^{pos}\right\| \tag{6}$$

where $\mathbf{x}_i^{YIQ}$ and $\mathbf{x}_i^{pos}$ indicate the color and the image coordinates of the reference patch $\mathbf{x}_i$, $\alpha_{col}$ and $\alpha_{pos}$ are normalization constants which make the color distance and coordinates distance between $\mathbf{x}_i$ and $\mathbf{x}_j$ within $[0, 1]$, and $C_{col}$ and $C_{pos}$ are weighting parameters.

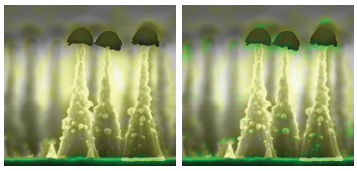

**Fig. 3:** Comparison between colorizing the Nano image with MRF inference (left) or not (right). MRF inference is important for producing smooth and consistent colorization results.

Inference in the MRF model is to minimize the energy function (5), and we adopt the fast graph cut method [13] for optimization. With the inferred optimal label $\{\mathbf{x}_i^*\}$, the missing color channels of $\mathbf{gI}$ are recovered by transferring color from $\{\mathbf{x}_i^*\}$ to the corresponding patches $\left\{\hat{\mathbf{Z}}_i\right\}$. The grayscale channel of $\mathbf{gI}$ is retained as the luminance channel after the color transfer process. To demonstrate the potential of MRF inference in rendering a smooth feature matching, we illustrate the result of colorizing the Nano mushroom-like images with and without MRF inference in Figure 3. Note that the Epitome-MRF model without MRF inference reduces to the feature matching scheme solely by $\mathcal{D}_{\mathbf{e}}$ in section 2.4.

## 3. EXPERIMENTAL RESULTS

We present the performance of epitomic image colorization with comparisons in this section. We learn the heterogeneous feature epitome of fixed size $[M_e, N_e, D]$ throughout the experiments, where $D = 12$ is the dimension of the heterogeneous feature and $M_e = N_e = 100$. The area of the heterogeneous feature epitome ($M_e \times N_e$) is no more than $\frac{1}{4}$ of that of the reference images. The patch size is $9 \times 9$ or $12 \times 12$. $\lambda_{col}, \lambda_{lbp}$ in (2) are set between $[0, 1]$, and we set $(\lambda_{col}, \lambda_{lbp}) = (0.1, 0.8)$ for the challenging Nano image, and use the fixed parameter $(\lambda_{col}, \lambda_{lbp}) = (0.8, 0.1)$ for all the other test images. The default setting for $C_{col}$ and $C_{pos}$ in (6) is $(1.0, 0.8)$.

We compare our method with the existing state-of-the-art automatic colorization methods, i.e. the color transfer method by Welsh et al. [3] and the most recent work by Gupta et al. [4]. Both are the representatives of automatic colorization

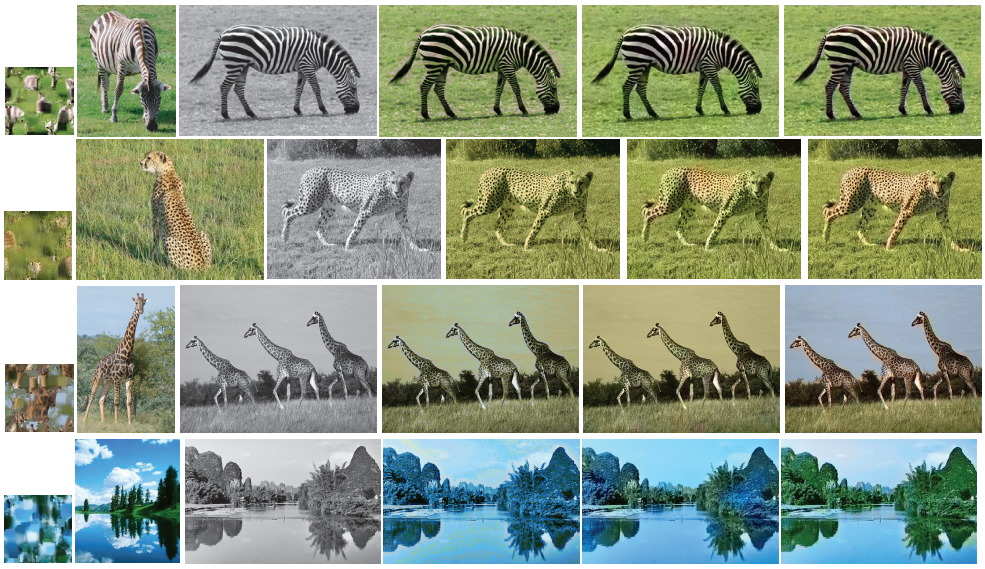

**Fig. 4:** Comparison with existing state-of-the-art automatic colorization methods by Welsh et al. and Gupta et al. For each row, from left to right: the heterogeneous feature epitome learned from the reference image, the reference image, the target image, the result by Welsh et al., the result by Gupta et al., and our result.

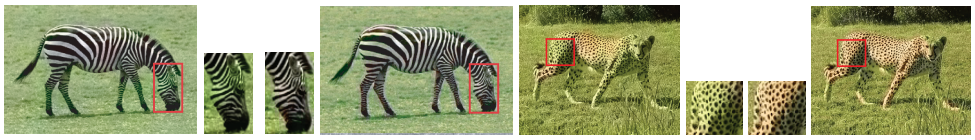

**Fig. 5:** Comparison between Gupta et al.'s method and our method on colorizing the marked region of the zebra and the cheetah.

methods, and we use the parameter settings suggested by the authors for both methods.

Figure 1 shows the colorization result for the Nano image containing mushroom-like structures acquired by scanning electron microscopy , where the reference image is colorized manually with Photoshop. It is a challenging colorization task since there is an "out of focus"effect for the body of the mushroom-like structures, and the bottom of the mushrooms are comprised of two strips of light green and dark green. Welsh's method adopts pixel-level matching based on luminance value and neighborhood statistics, so it suffers from spatial inconsistency which can be observed around the top of the mushroom structures. Gupta's method fails to colorize the bottom of the mushrooms correctly since its cascade feature matching scheme is not discriminative enough to find correct matching superpixels. Our method colorizes the mushroom-like structures faithfully with minimum artifacts.

Figure 4 compares the colorization results on several internet images. Welsh's method generates artifacts around the boundary of the zebra (the first row), and it colorizes the cheetah (the second row) with inaccurate pixel-level matching, resulting in a less appealing cheetah than that produced by Gupta's method and our method. The poses of the zebra and the cheetah undergoes large change across the reference and target image, which makes the cascade feature matching scheme [4] fail to accurately match all the target superpixels to the reference superpixels. In contrast, by matching the target patches to the epitome patches rather than the reference patches, our robust patch dissimilarity measure can still infer the reliable matching patches in the reference image by the generalization ability of the heterogeneous feature epitome. Our method also renders more visually appealing results on the giraffe, sky, trees and hill compared to the other two methods (the thrid and fourth row of Figure 4). Figure 5 demonstrates that our method produces more accurate feature matching for colorization.

## 4. CONCLUSION

We present an automatic colorization method called epitomic image colorization which transfers color from the reference color image to the target grayscale image. Our method employs the heterogeneous feature epitome to define a robust patch dissimilarity measure, and colorizes the target image by inference in the MRF model. Experimental results demonstrates the effectiveness of our method over other automatic colorization methods.

## 5. REFERENCES


[1] Anat Levin, Dani Lischinski, and Yair Weiss, "Colorization using optimization," *ACM Trans. Graph.*, vol. 23, no. 3, pp. 689–694, 2004.

[2] Qing Luan, Fang Wen, Daniel Cohen-Or, Lin Liang, Ying-Qing Xu, and Heung-Yeung Shum, "Natural image colorization," in *Rendering Techniques*, 2007, pp. 309–320.

[3] Tomihisa Welsh, Michael Ashikhmin, and Klaus Mueller, "Transferring color to greyscale images," *ACM Trans. Graph.*, vol. 21, no. 3, pp. 277–280, 2002.

[4] Raj Kumar Gupta, Alex Yong Sang Chia, Deepu Rajan, Ee Sin Ng, and Zhiyong Huang, "Image colorization using similar images," in *ACM Multimedia*, 2012, pp. 369–378.

[5] Nebojsa Jojic, Brendan J. Frey, and Anitha Kannan, "Epitomic analysis of appearance and shape," in *ICCV*, 2003, pp. 34–43.

[6] Chris Fraley and Adrian E. Raftery, "Model-Based Clustering, Discriminant Analysis, and Density Estimation," *Journal of the American Statistical Association*, vol. 97, no. 458, pp. 611–631, June 2002.

[7] Kai Ni, A. Kannan, A. Criminisi, and J. Winn, "Epitomic location recognition," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 31, no. 12, pp. 2158 –2167, dec. 2009.

[8] Xinqi Chu, Shuicheng Yan, Liyuan Li, Kap Luk Chan, and Thomas S. Huang, "Spatialized epitome and its applications," in *CVPR*, 2010, pp. 311–318.

[9] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the em algorithm," *Journal of the Royal Statistical Society: Series B*, vol. 39, no. 1, pp. 1–38, 1977.

[10] Christopher M. Bishop, *Pattern Recognition and Machine Learning (Information Science and Statistics)*, Springer-Verlag New York, Inc., Secaucus, NJ, USA, 2006.

[11] S. Lazebnik, C. Schmid, and J. Ponce, "Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories," in *Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on*, 2006, vol. 2, pp. 2169 – 2178.

[12] T. Ojala, M. Pietikainen, and T. Maenpaa, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 24, no. 7, pp. 971 –987, jul 2002.

[13] Yuri Boykov, Olga Veksler, and Ramin Zabih, "Efficient approximate energy minimization via graph cuts," *IEEE transactions on Pattern Analysis and Machine Intelligence*, vol. 20, no. 12, pp. 1222–1239, November 2001.