A Practical Approach for Super-Resolution using Photometric Stereo and Graph Cuts

Swati Sharma and Manjunath V. Joshi
Dhirubhai Ambani Institute of Information and Communication Technology
Gandhinagar, India
{sharma_swati, mv_joshi}@daiict.ac.in

Abstract

In this paper, we propose an approach to obtain super-resolved image and super-resolved depth map using photometric cue. The images are captured using different light source positions which are assumed to be known. The surface of the object is assumed to be Lambertian. We model the high resolution structure (surface gradients) as a Markov Random Field (MRF) and use graph cuts with discontinuity preservation to get a high resolution depth map. We then reconstruct the high resolution intensity map for each light source position using the high resolution surface gradients. Results of experimentation on synthetic and real data are presented. The advantage of the proposed approach is that its time complexity is much less as compared to the super-resolution approaches that use global optimization techniques such as simulated annealing. Also, since we are using photometric cue, there is no need of registration as is required in motion based approaches.

1 Introduction

Many existing vision applications require high spatial resolution images to take better decisions. Since the resolution of an image is dependent on the device which is used to acquire the image, it is difficult to use very high resolution sensors as they are often expensive. Hence, there is a need to develop efficient methods to obtain better quality high resolution images given the low resolution observations. Also, 3-D shape recovery of a scene is used extensively for applications such as object tracking and recognition. Super-resolution is the process of obtaining a high resolution image from several low resolution images of the same scene. Most researchers use motion cue to increase the resolution of an image. Although the 3-D structure of the scene being imaged is inherently available from the disparity map, the motion cue being a 2-D feature matching technique does not consider the 3-D structure. Hence, techniques to obtain high resolution images which preserve the structure are required [2]. This motivates us to look into the use of photometric cue in order to estimate the shape of the object and the high resolution image.

For practical vision applications, high resolution depth and intensity map estimation methods which are computationally efficient are required. However, since the problem is ill-posed, many researchers use regularization based approaches in order to obtain better estimates. Now, if the cost used for obtaining the solution is non-convex (discontinuity preserving cost), then optimization techniques such as simulated annealing are used to

BMVC 2007 doi:10.5244/C.21.83

obtain the global minima, which makes these methods very time consuming. For instance, if we consider an assembly line where an object has to be moved from one place to another (industrial inspection), the requirement is to be able to calculate depth fast enough so that the assembly line functions smoothly. Here the requirement is the speed and not very high accuracy. In such situations near global optimization methods such as graph cuts are useful.

In [8] the authors show the estimation of super-resolved image and depth map using photometric cue. They model the surface gradients and albedo as the Markov Random Field's (MRF) and use line fields for discontinuity preservation. They also use additional constraints for optimization. Since they use simulated annealing for minimization, the approach is computationally very taxing.

In this paper, we solve the problem of simultaneous estimation of the super-resolved depth map and intensity map using photometric cue. We use graph cuts for optimization which is much faster as compared to simulated annealing minimization approach even when a discontinuity preserving cost function is used. Our results show that the performance (both perceptually and quantitatively) of graph cuts based approach for super-resolution is better than general interpolation techniques. The results also show that our super-resolution approach takes much less time as compared to the approach using simulated annealing.

2 Previous Work

The idea of super-resolution was first proposed by Tsai and Huang [12]. In literature, many researchers have proposed approaches for super-resolution that use motion as a cue. In [4], Ur and Gross use the Papoulis-Brown generalized sampling theorem to obtain a high resolution image from several low resolution spatially shifted images. A set theoretic approach to the super-resolution restoration that is based on iterative back projection method adapted from computer-aided tomography was proposed in [6]. Here, the output image is initialized and the temporary results are projected to the measurements (by simulation). The temporary results are updated according to the simulation error. A regularized constrained total least squares based approach to obtain high resolution image was proposed in [7]. Cheeseman et al. [9] use a Bayesian method for reconstructing a super-resolved surface model by combining the information from a set of given images. They find the "emmitance" of the surface which is a combination of albedo, illumination conditions and ground slope for landsat images. In [14], the authors consider graph cuts optimization for super-resolution using motion cue. The high resolution image is modeled as MRF and graph cuts is used for optimization to get the super-resolved image.

Researchers have also explored the possibility of super-resolving the intensity map of the scene as well as the depth map. The authors in [5] formulate the problem of super-resolution depth reconstruction as that of expectation maximization and use a probabilistic approach using MRF modeling. In [3], Shekarforoush et al. use MRFs to model the images to obtain high resolution depth and albedo from a sequence of displaced low resolution observations. The effect of sampling a scene at a higher rate is acquired by having sub-pixel displacements.

In [13], graph cuts minimization technique has been used for estimation of the surface normals using photometric stereo. They use the ratio of two images, in order to cancel

out the albedo in the image irradiance equation, and get the initial estimates of the surface normal which are required to define the energy functions. Graph cuts is then used for optimization. Tan et al. proposed a technique in [10] for enhancing the resolution for photometric stereo. Their method first uses the generalized reflectance model to recover the distribution of surface normals inside each pixel, from which they infer sub-pixel surface geometry on a surface by spatially arranging the normals among pixels at a higher resolution.

3 High Resolution using Photometric Stereo

If a Lambertian surface is assumed, the image irradiance equation relating the surface gradients and image intensity can be written as,

$$E_l(x,y) = R(p_l(x,y), q_l(x,y)) = \rho_l(x,y)\hat{n}_l(x,y).\hat{s} \tag{1}$$

where $p_l(x,y), q_l(x,y)$ are the surface gradients in (x,y) directions respectively. Here $\rho_l(x,y)$ represents the albedo, which is nothing but the fraction of light reflected from the surface at the point (x,y) and its value lies between 0 and 1. $\hat{n}_l(x,y)$ denotes the surface normal given by $\frac{(-p_l(x,y),-q_l(x,y),1)}{\sqrt{p_l(x,y)^2+q_l(x,y)^2+1}}$ and $R(p_l(x,y), q_l(x,y))$ is the reflectance map, $E_l(x,y)$ is the image irradiance (or intensity) at point (x,y) in the image. $\hat{s} = \frac{(-p_s,-q_s,1)}{\sqrt{p_s^2+q_s^2+1}}$ is a unit vector in the direction of the light source. Here, the subscript l denotes low resolution.

It has been shown in [2] that generalized interpolation can be used with photometric stereo to obtain high resolution. The high resolution image can be reconstructed using the interpolated values of the surface gradients and albedo using Eq.(1). This technique is called generalized interpolation. The advantage of using photometric cue for obtaining high resolution observations is that since there is no relative motion between the scene and the camera, the need for image registration with sub-pixel accuracy is eliminated.

In this paper we use graph cuts optimization which considers the spatial dependency with discontinuity preservation. Our algorithm converges much faster than simulated annealing and hence can be applied in a practical scenario. It may be noted here that we do not optimize for albedo assuming that it is a smooth field and a simple interpolation method can be used to interpolate albedo, while combining high resolution surface gradients and albedo to get high resolution intensity image.

4 Proposed Approach for Super-resolution using Graph Cuts

Typically, in any reconstruction based super-resolution technique, the available information from a number of low resolution observations is used together to get a single super-resolved image. First, a forward model is defined to establish the low resolution image formation process which is then used to establish a relation between the desired high resolution image and the given low resolution images. On the basis of this relationship the high resolution image is then obtained using an inversion process. The inversion process being ill-posed, requires the use of regularization and a suitable optimization approach such as the one proposed here can be used to minimize the derived cost function to obtain

better estimates. It is shown in [8] that it is indeed possible to obtain super-resolution using low resolution observations captured at different source positions.

4.1 Image Formation Model

Let E_{l_m} be the vector containing the intensity values of the mth low resolution image of size $M \times N$ arranged in lexicographical order and of size $MN \times 1$, where, $m = 1 \ldots K$. K is the number of available images. Similarly, let $\hat{n}$ and ρ be the vectors that represent the *high resolution* surface normal and *high resolution* albedo arranged lexicographically. Now, if D is the decimation matrix which represents the aliasing due to under sampling, the low resolution image formation model can be expressed as,

$$E_{l_m} = DH\rho(\hat{n}.\hat{s}_m) + w_m \tag{2}$$

Here, $\hat{s}_m$ represents the light source position for the mth image. w_m is the independent and identically distributed (i.i.d) Gaussian distributed noise vector with variance σ_w^2. H represents the blurring matrix. In our implementation, we assume H as an identity matrix and we do not consider blurred observations. We choose the decimation matrix as the average of the corresponding pixels of the high resolution image as given in [11].

4.2 Cost Function Formation

Regularization is a popular method for solving computer vision problems which are ill-posed. The approach consists of minimizing a cost function which is a sum of two terms i.e. data fitting term and regularization term [1]. In order to form the regularization term, we model the surface gradients (p and q) as MRFs. With the image formation model expressed in Eq. (2), it is quite simple to write the cost function to be minimized for estimating the high resolution entities as,

$$\varepsilon = \sum_{m=1}^{K} ||E_{l_m} - D\rho(\hat{n}.\hat{s}_m)||^2 + \sum_{a=0}^{MN-1} [\lambda_p min(|p_a - p_b|, T_p) + \lambda_q min(|q_a - q_b|, T_q)] \tag{3}$$

The first term in the cost function is called the data cost that measures the deviation from the observed data, caused by assigning a particular label (here surface gradients in the x and y directions) to a pixel. The other two terms are discontinuity preserving MRF priors for two neighboring pixels a and b. Here, p_a and q_a represent the labels assigned to a pixel a. p and q are labels of the surface gradients in the x and y directions respectively. T_p and T_q are thresholds that are used for discontinuity preservation.

The data cost [first term of Eq. (3)] at pixel a_1 of the high resolution image is given as follows,

$$Data(a_1) = \sum_{m=1}^{K} (E_{l_m}([\frac{a_1}{r}]) - \frac{1}{r^2}(F(a_1) + \ldots + F(a_{r^2}))^2 \tag{4}$$

where $[.]$ represents the integer value. The function $F(a_1)$ represents the term $\rho(a_1)(\hat{n}(a_1).\hat{s}_m)$ for a particular pixel a_1. $F(a_1), F(a_2) \ldots, F(a_{r^2})$ are the pixel intensities of the r^2 pixels, $(a_1, a_2, \ldots, a_{r^2})$ of the high resolution image related to one pixel of the low resolution image according to the matrix D. For instance, if the up-sampling factor is 2, the pixel $(0,0)$ of the low resolution is related to the pixel locations $(0,0)$, $(0,1)$, $(1,0)$ and $(1,1)$ of the

high resolution image. Hence, the data cost for the pixel $(0,0)$ (and also for pixels $(0,1)$, $(1,0)$ and $(1,1)$) of the high resolution entity to be estimated is,

$$Data(0,0) = \sum_{m=1}^{K} (E_{l_m}(0,0) - \frac{1}{4}(F(0,0)+F(0,1)+F(1,0)+F(1,1)))^2 \quad (5)$$

In order to use the graph cuts formulation for optimization the cost function should be regular. Applications of graph cuts generally use the data term that is a function of a single pixel [15]. Thus, in order to apply the graph cuts formulation we use valid mathematical approximations. It can be observed from the cost function that image intensities of several pixels of the high resolution image are related to the image intensity of a single pixel in the low resolution image in the data cost. So, we treat the remaining r^2-1 terms $(F(a_2),F(a_3),\ldots,F(a_{r^2}))$ as constant for a particular optimization step. Then the modified data term can be written as follows,

$$Data(a_1) = \sum_{m=1}^{K} (E_{l_m}([\frac{a_1}{r}]) - \frac{1}{r^2}(F(a_1)+C))^2 \quad (6)$$

where $C = F(a_2)+\ldots+F(a_{r^2})$.

So, the modified total cost function at a pixel a can be expressed as,

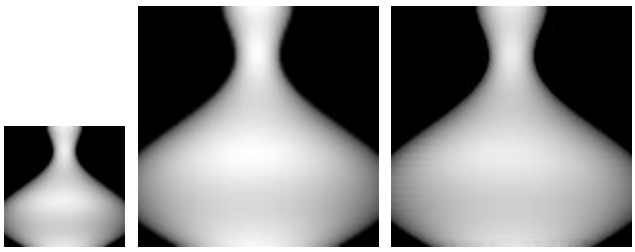
$$\varepsilon = \sum_{a=0}^{MN-1} [\sum_{m=1}^{K} (E_{l_m}([\frac{a}{r}]) - \frac{1}{r^2}(F(a)+C))^2 + \lambda_p.min(|p_a-p_b|,T_p)) + \lambda_q.min(|q_a-q_b|,T_q)] \quad (7)$$

where b is a neighboring pixel of a. The constant C represents the sum of the remaining r^2-1 terms, which are treated as constant for a particular optimization step.

We now optimize for the surface gradients, p and q using the graph cuts optimization. While optimizing for p field we consider q field as constant and vice versa. Both these fields are optimized one after another in each cycle until convergence is reached. The initial values for the high resolution p, q and ρ are obtained by interpolating the low resolution p, q and ρ fields (obtained using photometric stereo) by using a simple interpolation technique.

5 Choice of the Label Set

Graph cuts optimization requires a discrete label set. Most of the proposed methods that use graph cuts for optimization use integer labels. In our case, we use discrete floating point labels. Given the initial values of the surface gradients p and q, the range in which these fields lie is roughly known. Now based on the frequency distribution (histogram) of these labels it is possible to non-uniformly quantize the entire range of continuous values to get a discrete label set. The non-uniform quantization is done to assign maximum number of labels (discrete and integer) to that sub-range which has a higher probability. The number of labels, in this case, is directly related to the precision. As the chosen number of labels is increased, more accurate results may be obtained with a slight increase in computational complexity.



(a) (b) (c)

Figure 1: (a) Synthetically generated low resolution Vase image with light source position $(0,0,1)$ (b) Up-sampled image reconstructed using bi-cubic interpolation of p, q and ρ fields (c) Super-resolved Vase image using proposed approach

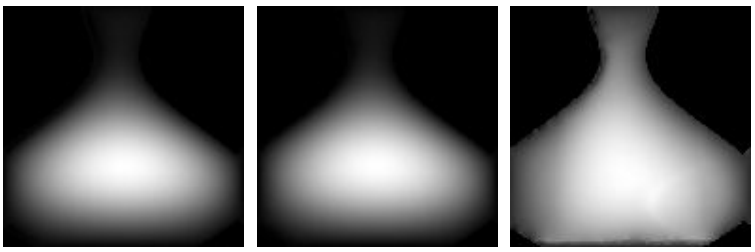
6 Experimental Results

In this section we present some of the results of our experiments. In order to test the performance of our algorithm, we first show results on a synthetic image Vase and then on a real image of a soft toy Jodu. We use the graph cuts library provided by Kolmogorov [18], [16], [17] with expansion algorithm for implementation.

First we consider the synthetic image Vase. Ten images of Vase of size 64×64 were generated using a computer program where each image is produced using a different light source position. These images are the given low resolution images. Now, in order to use graph cuts for optimization we need to use a fixed set of labels for each of the entities p and q. We observed that the initial values of p for the Vase image lie in the range $(-0.4, 0.6)$ and that of q lie in the range $(-0.2, 0.4)$. Hence, depending on the frequency distributions of the respective entities, we use 338 labels for p and 307 labels for q. The regularization parameters λ_p and λ_q for p and q respectively were manually adjusted to 0.01 and 0.01. The value of T of the truncated linear prior [See Eq. (3)] was chosen to be 0.8 for both p and q fields.

Fig. 1(a) shows the observed low resolution Vase image with light source position $(0,0,1)$ and of size 64×64. The Fig. 1(b-c) shows the images of size 128×128, reconstructed using bi-cubic interpolation of p, q and ρ fields and super-resolved image using the proposed method respectively. Although perceptually the images (b) and (c) look similar, the mean square error (MSE) comparison (discussed later) shows that graph cuts based approach is indeed better. Fig. 2(a) shows the high resolution ground truth for depth of Vase image. The Fig. 2(b-c) shows the up-sampled depth reconstructed using bi-cubic interpolation of p and q fields and super-resolved depth using the proposed method respectively. Perceptually the depth map obtained by using bi-cubic interpolation looks better than that obtained using the proposed method. However, by fine tuning the regularization parameter λ_x and the threshold T_x, where $x = p, q$, it is possible to get better results.

Next we consider a real object Jodu. Eight images of Jodu were captured with different light source positions. We consider the actual observed Jodu images of size 234×234 as the desired high resolution images. These images are decimated to obtain low resolution images of size 117×117, which now become the given low resolution observations.



(a) (b) (c)

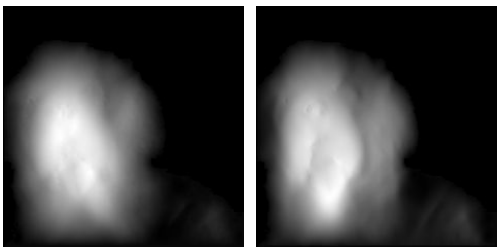
Figure 2: Depth map for Vase Image (a) Ground Truth (b) Up-sampled depth reconstructed using bi-cubic interpolation of p, q and ρ fields (c) Super-resolved using the proposed approach

(a) (b) (c)

Figure 3: (a) Observed image with light source position $(0.8389, 0.7193, 1)$ (b) Up-sampled image reconstructed using bi-cubic interpolation of p, q and ρ fields (c) Super-resolved Jodu image using proposed approach

For the Jodu image, we observed that the initial values of p lie in the range $(-1, 1)$ and that of q lie in the range $(-0.6, 0.6)$. Hence, depending on the frequency distributions of the respective entities, we use 440 labels for p and 420 labels for q. The regularization parameters λ_p and λ_q for p and q respectively were manually adjusted to 0.008 and 0.0259. The value of T of the truncated linear prior [See Eq. (3)] was chosen to be 0.175 for both p and q fields.

Fig. 3(a) shows one of the observed low resolution Jodu image of size 117×117. Fig. 3(b) shows the high resolution images of size 234×234 reconstructed from the bi-cubic interpolation of the p, q and ρ fields for the same light source positions. The super-resolved images using the proposed approach is shown in Fig. 3(c). Although visually there is not much difference in the super-resolved images reconstructed using graph cuts and bi-cubic interpolation, our quantitative analysis (discussed later) shows that the images reconstructed using graph cuts are indeed superior. The depth maps reconstructed using bi-cubic interpolation of p, q and ρ and that obtained using the proposed approach for the Jodu image are shown in Fig. 4(a-b). It may be noted here that we do not have the true depth map for comparison since the laser scanner does not work well with objects with discontinuities. One can observe from the Fig. 4(b) that discontinuities in depth



(a) (b)

Figure 4: Depth map for Jodu Image (a) Up-sampled depth reconstructed using bi-cubic interpolation p, q and ρ fields (b) Super-resolved using the proposed method

are much better revealed as compared to Fig. 4(a) that was reconstructed using bi-cubic interpolation.

For quantitative comparison, we use mean square error (MSE) as a figure of merit. Table 1 shows the MSE comparison for the super-resolved image and the depth map (for both Vase and Jodu images) and the case when interpolated values of the surface gradients and albedo are used for reconstruction of the up-sampled depth and intensity map. Although, not much difference can be seen in the high resolution images reconstructed using the two methods, the MSE values clearly show that the high resolution images obtained using our graph cuts based approach are much better than those obtained using bi-cubic interpolation. Due to the use of edge preserving smoothness term, the reconstructed image using graph cuts minimization is closer to the actual high resolution images. The high resolution depth obtained for the Vase image using our approach shows a superior MSE performance as compared to bi-cubic interpolation. It may be mentioned here that we do not have the actual depth map for Jodu, we use the depth map obtained using the actual observed 234×234 images with photometric stereo as the reference depth map for calculating MSE. Since the reference depth map is not the actual depth map (with edges properly defined), the MSE performance when depth is obtained using the proposed approach is poorer when compared to depth obtained using bi-cubic interpolation.

We now discuss the time complexity of our algorithm. The graph cuts based super-resolution approach takes around $5-7$ minutes for convergence (on a 1.33 GHz processor for 234×234 image size) while it takes hours for convergence when simulated annealing with edge preservation is used [8]. In [8] the authors mention that the time for convergence using simulated annealing is of the order of hours. Our approach, on the other hand, takes few minutes. It may be mentioned that although we are not using the other constraints used in [8] while optimization since the time required for simulated annealing is much larger as the cost is computed by changing the label of a single pixel in each move. On the other hand, in graph cuts the labels of a number of pixels get changed together in each move. One can thus observe the kind of complexity reduction that has been achieved through the graph cuts based formulation for super-resolution. Hence, our approach performs much better when compared to computationally expensive optimization methods. It may also be mentioned here that in [8] the discontinuity preservation prior terms consisted of edge preserving line fields. However, we use a truncated absolute distance for edge preservation. Hence, we do not compare our results with the simulated

Table 1: MSE comparison for the high resolution Vase and Jodu images and depth map obtained using bi-cubic interpolation and our super-resolution approach with an upsampling factor of 2 with different source positions. The (DEPTH) row in the table gives the MSE for the depth field.

Source position for Vase Image	MSE	
	Bi-cubic Interpolation	Graph cuts
(0, 0, 1)	86.03	14.82
(DEPTH)*	6.71	1.68
For Jodu Image		
(0.8389, 0.7193, 1)	240.13	43.55
(-0.1763, -0.5596, 1)	544.01	51.00
(DEPTH)	9.57	68.79

* Only the center portion of the Vase is used for MSE calculation.

annealing based super-resolution method proposed in [8].

7 Conclusion

In this paper, we used graph cuts optimization for obtaining a super-resolved depth map and intensity map using photometric cue. The surface gradients were modeled as separate MRFs. We used a smoothness prior with discontinuity preservation. The results show that the super-resolved image and depth obtained using our approach reveal edges better than the up-sampled depth and images obtained using general interpolation techniques. The quantitative measure (MSE) also shows that the graph cuts based super-resolution scheme is superior than these methods. Also, our approach takes a few minutes for convergence which is very much less than the super-resolution scheme that uses simulated annealing for optimization [8] (takes hours for convergence). It can be seen from the results that our graph cuts based super-resolution approach provides a time-effective method for super-resolution which is very much required in a practical scenario.

8 Acknowledgements

We are thankful to Dr. André Jalobeanu, LSIIT, Université de Louis Pasteur, Strasbourg, France for his constructive suggestions and comments.

References

[1] Tikhonov A. N. and Arsenin V. Y. *Solution of Illposed Problems*. W.H. Winston, Washington D.C., 1977.

[2] Rajan D. and Chaudhuri S. Generalized interpolation and its application in super-resolution imaging. *Image and Vision Computing*, 19(13):957–969, 2001.

[3] Shekarforoush H., Berthod M., Zerubia J., and Verman M. Sub-pixel bayesian estimation of albedo and height. *International Journal of Computer Vision*, 19(3):289–300, 1996.

[4] Ur H. and Gross D. Improved resolution from sub-pixel shifted pictures. *CVGIP: Graph, Models and Image Process.*, 54, 1992.

[5] Berthod M., Shekarforoush H., Verman M., and Zerubia J. Reconstruction of high resolution 3d visual information. *Technical Report, RR-2142, INRIA*, 1993.

[6] Irani M. and Peleg S. Improved resolution by image registration. *CVGIP: Graph, Models and Image Process.*, 53, 1991.

[7] Ng M. K., Koo J., and Bose N. K. Constrained total least squares computation for high resolution image reconstruction with mulisensors. *International Journal of Systems and Technologies*, 12, 2002.

[8] Joshi M. V. and Chaudhuri S. Simultaneous estimation of super-resolved depth map and intensity field using photometric cue. *Computer Vision and Image understanding*, 101:31–44, 2006.

[9] Cheeseman P., Kanefsky B., Hanson R., and Stutz J. Super-resolved surface reconstruction from multiple images. *Technical Report, FIA-94-12, NASA Ames Research center, Artificial Intelligence Branch*, 1994.

[10] Tan P., Lin S., and Quan L. Resolution-enhanced photometric stereo. *European Conference on Computer Vision*, 3:58–71, 2006.

[11] Schultz R. R. and Stevenson R. L. A bayesian approach to image expansion for improved definition. *IEEE Transactions on Image Processing*, 3(3):233–242, 1994.

[12] Tsai R. Y. and Huang T. S. Multiframe image restoration and registration. *Advances in Computer Vision and Image Processing, JAI Press*, 1984.

[13] Wu T. P. and Tang C. K. Dense photometric stereo using a mirror sphere and graph cut. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2005.

[14] Mudenagudi U., Singla R., Kalra P., and Banerjee S. Super-resolution using graph-cuts. *Asian Conference on Computer Vision*, 2006.

[15] Kolmogorov V. and Zabih R. Multi-camera scene reconstruction via graph cuts. *European Conference on Computer Vision*, 3:82–96, 2002.

[16] Kolmogorov V. and Zabih R. What energy functions can be minimized via graph cuts? *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26, 2004.

[17] Boykov Y., Veksler O., and Zabih R. Fast approximate energy minimization via graph cuts. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23:1222–1239, 2001.

[18] Boykov Y. and Kolmogorov V. An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2004.