Supervised Geodesic Propagation for Semantic Label Transfer

Xiaowu Chen*, Qing Li*, Yafei Song, Xin Jin, and Qinping Zhao

State Key Laboratory of Virtual Reality Technology and Systems
School of Computer Science and Engineering, Beihang University, Beijing, China
{chen,liqing,songyf,jinxin,zhaoqp}@vrlab.buaa.edu.cn

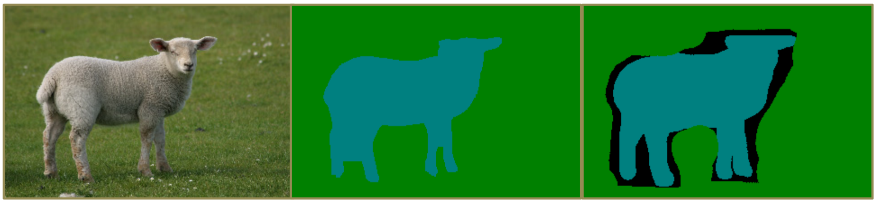
Abstract. In this paper we propose a novel semantic label transfer method using supervised geodesic propagation (SGP). We use supervised learning to guide the seed selection and the label propagation. Given an input image, we first retrieve its similar image set from annotated databases. A Joint Boost model is learned on the similar image set of the input image. Then the recognition proposal map of the input image is inferred by this learned model. The initial distance map is defined by the proposal map: the higher probability, the smaller distance. In each iteration step of the geodesic propagation, the seed is selected as the one with the smallest distance from the undetermined superpixels. We learn a classifier as an indicator to indicate whether to propagate labels between two neighboring superpixels. The training samples of the indicator are annotated neighboring pairs from the similar image set. The geodesic distances of its neighbors are updated according to the combination of the texture and boundary features and the indication value. Experiments on three datasets show that our method outperforms the traditional learning based methods and the previous label transfer method for the semantic segmentation work.

1 Introduction

Semantic labeling, or multi-class segmentation, is a basic and important topic in computer vision and image understanding. In the last decades, there has been a great advance in this research area [1–5]. However it is still a challenging problem for current computer vision technique to recognize and segment objects in an image as human beings. Recently, some methods have been proposed to typically solve this problem with trained generative or discriminative models [1–3] which usually need a fixed dataset contains certain category classes. Moreover, some approaches aim to integrate the low level features and context priori into the bottom-up and top-down model [4]. These methods require the training procedure on the fixed dataset for the parametric model and they are not scalable with the number of object categories. For example, to include more object category for the learning-based model, we need to train the model with additional label categories to adapt the model parameters.

* Corresponding authors.

A. Fitzgibbon et al. (Eds.): ECCV 2012, Part III, LNCS 7574, pp. 553–565, 2012.



(a) Input image (b) Our result (c) Ground truth

Fig. 1. The objective. Our objective is to get the semantic segmentation of the input image. This input image is taken from MSRC dataset[6].

With the increasing availability of image collections, such as LabelMe database [7], large data driven methods have revealed the potential for nonparametric methods in several applications, such as object and scene recognition [8] and semantic segmentation [9, 10]. Since Liu *et al.* [9] addressed the nonparametric scene parsing work as *label transfer* for the first time, several scholars have paid attention on this topic [10–12] and given promising results. Label transfer, considered as transfer the label of existing annotation to the unlabeled input image, involves two key issues to be solved. The first issue is how to retrieve proper similar images from the database for a given input image. The second issue is how to parse the input image with the annotated similar images. The first issue is well studied in some previous works [8, 13] and is not the main focus of our work. To solve the second issue, it needs a precise matching between the similar images and the input image, which is the dominate point in [9–12].

As above label transfer methods usually use MRF optimization followed by the matching, complete pixel-level or superpixel-level matching between similar images and input image needed to be implemented. We thus take the idea of partly matching on a little number of superpixels and selecting the initial seeds to propagate the semantic label. Some segmentation works [14–16] mainly grow regions based on the foreground/background seed employing the geodesic distance metric. In addition, Chen *et al.* [17] employ the geodesic propagation in multi-class segmentation. Inspired by these works, we present a novel method for label transfer using supervised geodesic propagation (SGP).

Given an input image, we first retrieve its K-Nearest-Neighbor (KNN) similar images to form its similar image set from annotated database with GIST matching [18]. Then the proposal map of the input image is inferred by the boosted recognition model learned on its similar image set. Besides the proposal map, a classifier for propagation indication is also learned on this similar image set. As the initial distance map of input image has been derived from the proposal map, we start our geodesic propagation guided by selected seed and the indication of label propagation. Moreover, the texture and boundary features of input image are also integrated into the propagation procedure.

Our main contributions include: (1)A label transfer method with geodesic propagation. (2)Supervised seed selection and label propagation scheme in geodesic propagation for label transfer. Figure 1 shows our label transfer result.

2 Related Works

Liu *et al.* [9] firstly address the semantic labeling work into a nonparametric scene parsing framework denoted as *label transfer*. Following the idea of label transfer, Tighe and Lazebnik [10] use scene-level matching with global features, superpixel-level matching with local features and Markov random field (MRF) optimization for incorporating neighborhood context. However, these two non-parametric methods both require an existing large training database. In addition, finding highly similar training images for a given query image is difficult. To assure that the retrieved images are proper, multiple image sets cover all semantic categories in the input image are firstly obtained in [11]. Then a KNN-MRF matching scheme is proposed to establish dense correspondence between the input image and each found image sets. A Markov random field optimization is used based on those matching correspondences. In [12], a ANN bilateral matching scheme is proposed. Both the retrieval of partially similar images and the parsing scheme are built upon the ANN bilateral matching. The test image is parsed by integrating multiple cues in a markov random field. Inspired by these works, our method take the similarity of retrieved images for semantic labeling without precise pixel-level or super-pixel level matching.

Geodesic distance which is the shortest path between points in the space is used as a metric to classify the pixels in [14]. The method proposed in [15] is also based on the idea of seed growing geodesic segmentation. To avoid the bias to seed placement and the lack of edge modeling in geodesic, Price *et al.* [15] present to combine geodesic distance information with edge information in a graph cut optimization framework. Gulshany *et al.* [16] introduce Geodesic Forests, which exploit the structure of shortest paths in implementing the star-convexity constraints. In addition to the methods [14–16] that mainly focus on the foreground/background segmentation with geodesic distance, Chen *et al.* [17] employ geodesic distance to multi-class segmentation by integrating color and edge constraints on the edge weight. However, they [17] utilize no contextual similarity in the propagation. Besides, we are obviously different from [17] in the training set and the seed selection scheme. We are somehow similar with [19, 20] in using GIST feature and boosting classifier, however, we use geodesic propagation to get the deterministic solution while [19, 20] use CRF or MRF to get the optimal solution. In addition, we utilize the contextual information of both the similar images and the test image itself, while [19, 20] utilize only the contextual information of the test image or test sequence.

3 Supervised Geodesic Propagation

The workflow of our method is illustrated in Figure 2. Given an input image, our method begins with getting the similar image set from the annotated dataset

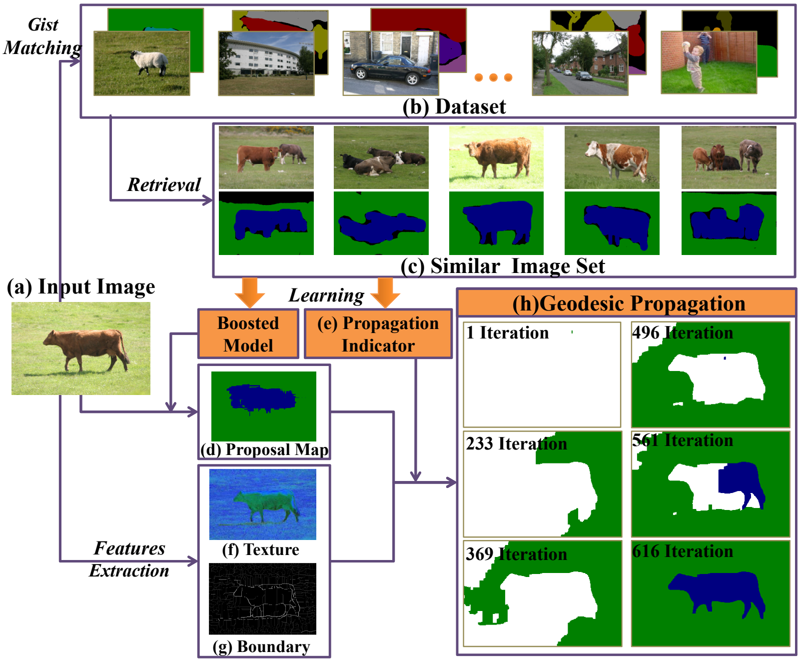


Fig. 2. Overview of our supervised geodesic propagation. (a) The input image. (b) The dataset with manually annotations. Each color uniquely maps to a category label. The semantic label maps are best viewed in color. (c) We firstly get the similar image set from the dataset with Gist matching and re-ranking (Section 3.1). (d) Then we infer the proposal map of the input image using the boosted model which is trained on this similar image set (Section 3.2). (e) The propagation indicator is learned on the similar image set (Section 3.3). (f) and (g) are the texture and boundary features of the input image (Section 3.4). (h) The propagation indicator, the proposal map, the texture and boundary features of the input image are integrated into the geodesic propagation framework (Section 3.4) to get the semantic label result of the input image.

through Gist matching [18]. We infer the proposal map of the input image for seed selection. Then the proposal map, the texture and boundary features of input image, and the contextual similarity of the similar images are integrated into the geodesic propagation framework to get segmentation result. There is no explicit energy to minimize in the framework, instead, multi-class semantic labeling can be solved through geodesic propagation process.

The denotations in this paper are presented here for clarity. Given an input image I, the semantic labeling problem is to assign each pixel $x \in I$ with a certain label $l \in L; L = \{1, 2, \cdots, N\}$. To solve this problem, we build up a neighbor-connected graph $G =< V, E >$. In this paper, each superpixel i is a vertex v in graph G and is assigned a specific label l contained in the dataset through geodesic propagation procedure. The edge set E consists of edges between neighboring vertices. We define the weight of edge $W(i, j)$ on a hybrid manifold based on both texture and boundary feature to indicate the smoothness between neighboring vertices i and j.

3.1 Similar Image Matching

In order to transfer labels of annotated images to the input image, the first thing we need to do is to select proper images similar to the input image in semantic categories and contextual informations. Since the similar image retrieval is not the main focus of this paper, we use Gist matching as recent label transfer methods [9–12] did. The gist descriptor [18] is employed to retrieve the K-Nearest Neighbors from the dataset and the similar image set is formed with these neighbors. After gist matching, the K-Nearest neighbors in the similar image set are re-ranked in the following way. We over-segment the input image and each of its similar image $R \in \{R\}$ using algorithm described in [21]. Then each superpixel $i \in I$ is matched to a proper superpixel $r(i) \in R$ which has the smallest matching distance to i. The following distance metric is used to compute the matching distance of correspondence in the re-ranking procedure. Given two images I and R, the matching distance $D_r(I, R)$ is scored as:

$$D_r(I, R) = \sum_{i \in I, r(i) \in R} \left\| (fv_i - fv_{r(i)}) \right\|^2 \tag{1}$$

Here, fv_i is a 22 dimension descriptor of i, including average HSV colors, coordinates and 17 dimension filter responses [6] of i. The Euclidean distance metric is used in our implementation. After re-ranking the gist similar images according to their matching scores, we get the top K similar images which are denoted as $\{R_K\}$. In our experiments, we use the $\{R_K\}$ as the similar image set instead of $\{R\}$.

3.2 Proposal Map for Seed Selection

Previous works [14–16] utilize the manual scribbles as the initial seeds for each category or objects which take the human priori into the segmentation task. Instead, we take a dynamic seed selection for geodesic propagation. The similar images in $\{R_K\}$ imply the possible categories in the input image. We assume that categories $l \in \{R_K\}$ cover all the categories in I. To exploit the inherent possibilities, the similar image set R_K is used as the training set for the input image I to learn the recognition model. The 17 dimension raw texton features and Jointboost algorithm are used in the learning procedure as [22] did.

When we get the recognition proposal map of I, the initial geodesic distances of all classes for each superpixel are defined upon this proposal map. Each superpixel is temporarily assigned the initial label that has the max probability $p_l(i)$. According to equation 2, we get the initial geodesic distance of their temporary class for each superpixel. Then a distance map of the input image can be obtained: superpixels which have higher probabilities will have smaller initial geodesic distance. In each propagation step, the undetermined superpixel with the smallest distance of all classes is selected as the current seed to ensure the estimation is the best in probability (see Section 3.4 for more details) .

$$Dis_{initial}(i) = 1 - p_l(i) \tag{2}$$

3.3 Propagation Indicator

Each image $R \in \{R_K\}$ is similar with the input image in some aspects, such as the appearance and the contextual information. We thus assume that the contextual similarity between the similar image set and the input image can provide useful information for parsing the input image. Based on this assumption, we take a supervised classification for label propagation. A set of classifiers is learned on the similar image set to guide the propagation and each semantic category has its corresponding classifier. We denote these classifiers as the propagation indicators. In this section, we introduce how to get the indicator of every semantic category. More details about how these indicators work in the propagation procedure will be introduced in section 3.4.

Our indicator is used to classify whether to propagate label from superpixel i to its neighbor j in the input image. For neighboring superpixels which are classified as the same category, we propagate current label; otherwise, we do not propagate current label. Our propagation indicators for each category are trained using random forests [23, 24], a competitive non-linear model that predicts by averaging over multiple regression trees. The random forests implementation available online [25] is used with default parameters in our method.

To generate training samples for the indicator of each category l in $\{R_K\}$, we get all neighboring superpixel pairs (i, j) as well as their category labels l_i and l_j known from the annotation. Note that pair (i, j) is different from pair (j, i). For each pair (i, j), we denote $fv(i, j) =< fv_i, fv_j >$ as a 44 dimension feature vector, which includes average HSV colors, coordinates and 17 dimension filter responses [22] of both fv_i and fv_j. If l_j is consistent with l_i, then $fv(i, j)$ is taken as a positive sample of the indicator of label l_i, otherwise a negative sample. All the features in fv are normalized in the range of $[0, 1]$. In the testing procedure, we extract the $fv(i, j)$ feature vector of neighboring superpixels and then get the confidence exported by trained classifier as an indicator value for propagation. As shown in equation 3, $T_l(i, j)$ is the indicator function, $con_l(i, j)$ is the confidence and φ is the threshold for indicator.

$$T_l(i, j) = 1[con_l(i, j) > \varphi] \tag{3}$$

3.4 Geodesic Propagation

We take the definition that geodesic distance is the smallest integral of a weight function over all possible paths from the seeds to a point [14, 17]. There is no explicit energy to be minimized in our propagation. Our objective is to assign each pixel the category label which has the smallest geodesic distance. The weight W

of edge E on the graph G denote the smoothness relationship between neighboring superpixels. There are many features can be included in the weights. To propagate correct labels to superpixels, we exploit intra contextual information of the input image by using color, texture and boundary features. These features can indicate the real object probability distribution on the input image to a certain extent [6, 21]. In this section, we will introduce our geodesic propagation procedure which simultaneously propagates geodesic distance of all the classes efficiently.

Here we integrate two components to measure the weight of edge W_{ij} on graph G: the texture component and the boundary component. The weight function W between neighboring vertices is demonstrated in equation 4, where λ_1 and λ_2 are tuning parameters. Regions of different categories can commonly present apparent texture disparities. We measure the $W_{texture}(i, j)$ with Euclidean distance metric using a texture descriptor consist of average HSV colors and 17 filter responses features [22]. The reliable Berkeley edge detector [21] is applied combining color, brightness and texture cues to capture the boundary confidence. The weight function for boundary component W_{bdry} is defined in equation 5, in which θ is the threshold for boundary confidence $P_b(\cdot)$. We detect the boundaries in pixel level and then convert these boundary confidences into superpixel level.

$$W(i, j) = \lambda_1 W_{texture}(i, j) + \lambda_2 W_{bdry}(i, j) \tag{4}$$

$$W_{bdry}(i, j) = P_b(i, j, \theta) \tag{5}$$

Our supervised geodesic propagation algorithm starts with the initial geodesic distance and initial labels for all the vertices. In each step of propagation, vertices which have undetermined labels will be put into the unlabeled set Q to sort for current seed that has minimum geodesic distance. The initial seed for propagation is obtained by the recognition proposal map. Once a vertex is selected as seed in a step, it is removed out of the set Q with its semantic label determined as current label. The weight of edge $W(i, j)$ and propagation indicator are integrated in the propagation iteration to decide how to update geodesic distance. In each step of geodesic propagation, the geodesic distance between a labeled seed and its neighboring undetermined superpixels have to be updated according to corresponding indicator. Suppose i has labeled as l_i and j has undetermined label, then employ the propagation indicator of category l_i and get the indicator confidence value $T_l(i, j)$. If $T_l(i, j)$ is true and $W(i, j)$ is less than threshold θ_e, then the geodesic distance $Dis(j)$ of j is updated as $Dis(i) + \kappa W(i, j)$ and the label of j is assigned l_i; otherwise $Dis(j)$ and the label of j stay their current statuses. Figure 2 (h) illustrates the propagation process step by step. The semantic labels are propagated from initial seed to the entire image after iterating 616 times. Comparing the 233, 369, 496 and 561 iteration, we can see that our method work well around the object edges and our algorithm propagates semantic label among multiple classes. The algorithm is convergent in a linear time depending on the image size and we implement a queue introduced in [26].

(a) Seed Selection (b) Before Update Neighbors (c) After Update Neighbors (d) Final Result

Fig. 3. Illustration of one iteration. Take the three statuses of 313 iteration for details instruction. Dark green means superpixels determinately labeled as the class *grass* and dark blue determinately the *cow*. Light green means superpixels temporarily labeled as the class *grass* and light blue temporarily the *cow*. (a) In the first status, the current seed is selected (in red curve) and its label is determined as *grass*. (b) Then in the second status, we get the undetermined neighboring superpixels of current seed for their update of both label and geodesic distance. We only shows parts of its neighbors for examples. The intra features of the input image and our indicator jointly make decision that whether to update the current label and distance of these neighhors. (c) In the third status, one neighbor is updated while the other is not. Then it is ready for the next iteration. (d) In the final iteration (616), we get the final result with all superpixels have their determined labels.

To illustrate how our supervised indicator guide the propagation direction, we choose one step in the propagation as shown in Figure 3. Dark green demonstrates superpixels determinately labeled as the class *grass* and dark blue determinately the *cow*. Light green demonstrates superpixels temporarily labeled as the class *grass* and light blue temporarily the *cow*. The first status is the beginning of the 313 iteration: the current seed which is denoted in red curve is selected with its label being determined as *grass*. Then in the second status, the undetermined neighboring superpixels of current seed is figure out for their update. The intra features of the input image and our indicator jointly make decision that whether to update the current geodesic distances and labels of these neighhors. Note that we show the 313 iteration for instance and the 314-615 iterations are omitted. The final iteration (616) is displayed for easy comparison.

4 Experiments

In this section we investigate the performance of our method on several challenging datasets, and compare our results with several state-of-the-art works. The datasets are the Cambridge-driving Labeled Video dataset (CamVid) [27], the Microsoft Research Cambridge (MSRC) dataset [6] and the CBCL StreetScenes (CBCL) dataset [28]. Table 1 shows our semantic segmentation accuracy compared with other methods. *Our Method without GP* indicates the accuracy of the proposal map (without the Geodesic Propagation). *Our Method* indicates the accuracy after our geodesic propagation procedure. We are the best on the CamVid dataset and the MSRC dataset. Although our method is 1.1% lower than [11] on the CBCL dataset, we are 3.46% higher than [11] on the CamVid dataset. Moreover, our method performs better than [3] significantly on the CBCL dataset.

Table 1. Comparison of semantic segmentation accuracy over three datasets

Method	CamVid dataset	MSRC dataset	CBCL dataset
Zhang *et al.* [11]	84.4%	-	**72.8%**
Shotton *et al.* [22]	-	72.2%	-
Tu [1]	-	77.7%	-
Shotton *et al.* [3]	-	-	61.9%
Our Method without GP	86.09%	75.4%	67.7%
Our Method	**87.76%**	**79.2%**	71.7%

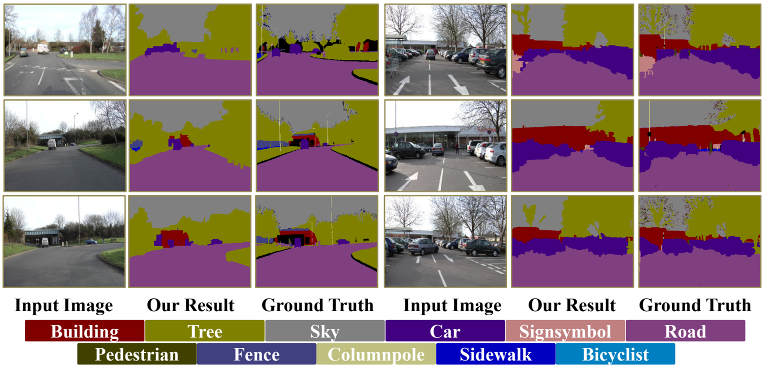


Fig. 4. Our results on CamVid dataset. Left column shows the input image and middle column shows our label transfer result. Ground truth is shown on the right. Legend is shown on the bottom.

In our experiments, the training procedure of Joint Boost model takes about 40 seconds per image and the training of label propagation scheme for each image is also about 40 seconds. Geodesic propagation takes about 5 seconds. On each dataset, we set 500 rounds to train the Joint Boost model for each test image. All the experiments are implemented on PC computers. Some results on the three datasets are shown in Figure 4, Figure 5 and Figure 6.

4.1 CamVid Dataset

The CamVid dataset is the first collection of videos with object class semantic labels. It provides 701 still images taken under different lighting condition (day and dusk). The images in the original dataset are at the size of 960× 720 and cover 32 object classes. To compare with others, we group the dataset into 11 categories as [11] did and resize the images to 480×360 pixels. The 11 dominant categories are building, tree, sky, car, sign-symbol, road, pedestrian, fence, column-pole, sidewalk, bicyclist. Besides, a *void* labeling indicates that the pixel does not belong to the 11 categories. We use 5 similar images for each test image

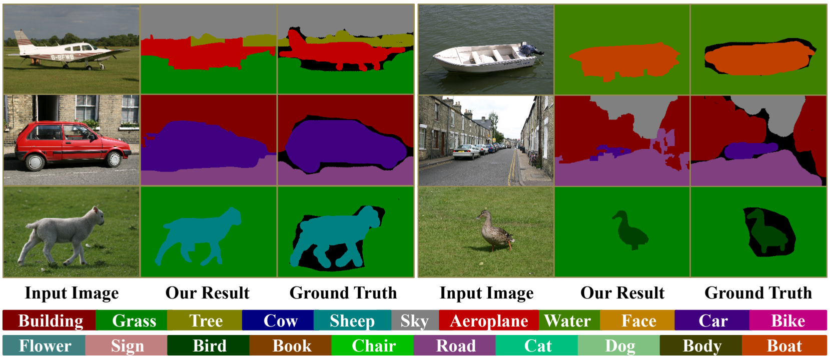


Fig. 5. Our results on MSRC dataset. Left column shows the input image and middle column shows our label transfer result. Ground truth is shown on the right. Legend is shown on the bottom.

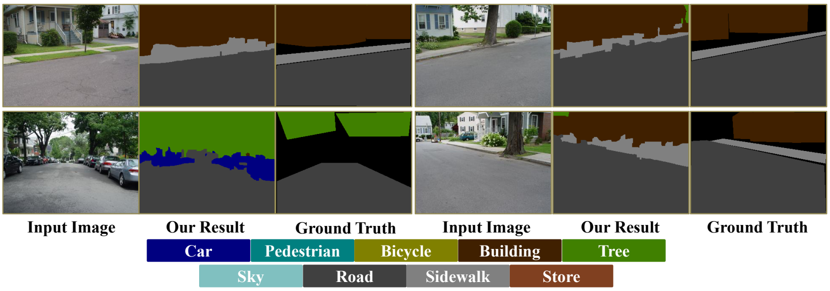


Fig. 6. Our results on CBCL dataset. Left column shows the input image and middle column shows our label transfer result. Ground truth is shown on the right. Legend is shown on the bottom.

to train the propagation indicator. The threshold θ_e and φ are set to be 0.8 and 0.75 respectively. The segmentation accuracy of our method is 87.76% as shown in Table 1.

4.2 MSRC Dataset

MSRC dataset is composed of 591 images of 21 object classes. We randomly split this dataset into 491 training images and 100 testing images. The *void* labeling is used to cope with pixels not belonging to a class in the dataset, and the manual labeling is not aligned exactly with boundaries. Image in this dataset is about 320×213 pixels. We use 10 similar images for each test image to train the propagation indicator. The threshold θ_e and φ are set to be 0.5 and 0.6 respectively. The segmentation accuracy of our method on this dataset is 79.2%.

Figure 5 shows some results of proposed method on this dataset. The comparisons with previous methods are shown in Table 1. Our method is the best among those methods on this dataset.

4.3 CBCL Dataset

The CBCL StreetScenes dataset contains 3547 still images of street scenes, which includes nine categories: car, pedestrian, bicycle, building, tree, sky, road, sidewalk, and store. In our test, the pedestrian, bicycle, and store are not included as [11] did. To compare with [11], we resize the original images to 320×240 pixels. We use 5 similar images for each test image to train the propagation indicator. The threshold θ_e and φ are set to be 0.3 and 0.6 respectively. The segmentation accuracy of our method is 71.7% as shown in Table 1. Although our method is not the best on this dataset (about 1 percent lower than [11]), we perform better than [3] who has accuracy of 61.9%. Figure 6 shows the results of our method.

5 Conclusion and Discussion

In this paper we propose a supervised geodesic propagation for semantic label transfer. Following the label transfer idea, given an input image, we first retrieve its top K-Nearest-Neighbor similar images from annotated dataset with GIST matching. We then infer the recognition proposal map of the input image with the similar images based on the trained model. The similar image set is not only used to learn the category proposals, but also used to train the label propagation scheme. Unlike other label transfer methods which focus on matching the test image with the similar images, we employ the supervised geodesic propagation to utilize the similar images for semantic labeling of input image. Then we begin geodesic propagation that starts from dynamic selected seeds. Experiments on three datasets show that our method outperforms the traditional learning based methods and the previous label transfer methods for the semantic segmentation work.

Limitations and Future Work. As we consider the similar image set cover all categories in the input image, our method is sensitive to the retrieved similar images. If the retrieved images have little similarity with the test image, or the retrieved images do not have the category in the test at all, our method will fail. In future, we will pay attention on how to retrieve similar image set of high quality. In addition, some parameters in our implementation are manually selected. Next we will focus on adaptive parameter selection scheme.

Acknowledgments. In addition to the anonymous reviewers who provided insightful suggestions, the authors would like to thank Yi Liu for her data processing. This work was partially supported by NSFC (60933006), 863 Program (2012AA011504), R&D Program (2012BAH07B01), Special Program on ITER (2012GB102008) and BUAA (YWF-12-LKGY-001 and YWF-12-RBYJ-035).

References

1. Tu, Z.: Auto-context and its application to high-level vision tasks. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 1–8. IEEE Press, Piscataway (2008)
2. Gould, S., Fulton, R., Koller, D.: Decomposing a Scene into Geometric and Semantically Consistent Regions. In: IEEE International Conference on Computer Vision, pp. 1–8. IEEE Press, Piscataway (2009)
3. Shotton, J., Johnson, M., Cipolla, R.: Semantic texton forests for image categorization and segmentation. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 1–8. IEEE Press, Piscataway (2008)
4. Han, F., Zhu, S.-C.: Bottom-up/Top-Down Image Parsing by Attribute Graph Grammar. In: IEEE International Conference on Computer Vision, pp. 1778–1785. IEEE Press, Piscataway (2005)
5. Zhao, P., Fang, T., Xiao, J., Zhang, H., Zhao, Q., Quan, L.: Rectilinear parsing of architecture in urban environment. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 342–349. IEEE Press, Piscataway (2010)
6. Shotton, J., Winn, J., Rother, C., Criminisi, A.: TextonBoost for Image Understanding: Multi-Class Object Recognition and Segmentation by Jointly Modeling Texture, Layout, and Context. Int. J. Comput. Vis. 81, 2–23 (2009)
7. Russell, B., Torralba, A., Murphy, K., Freeman, W.: Labelme: A database and web-based tool for image annotation. MIT AI Lab Memo (2005)
8. Torralba, A., Fergus, R., Freeman, W.: 80 Million Tiny Images: A Large Data Set for Nonparametric Object and Scene Recognition. IEEE Trans. Pattern Anal. Mach. Intell. 30, 1958–1970 (2008)
9. Liu, C., Yuen, J., Torralba, A.: Nonparametric scene parsing: Label transfer via dense scene alignment. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 1972–1979. IEEE Press, Piscataway (2009)
10. Tighe, J., Lazebnik, S.: SuperParsing: Scalable Nonparametric Image Parsing with Superpixels. In: Daniilidis, K., Maragos, P., Paragios, N. (eds.) ECCV 2010, Part V. LNCS, vol. 6315, pp. 352–365. Springer, Heidelberg (2010)
11. Zhang, H., Xiao, J., Quan, L.: Supervised Label Transfer for Semantic Segmentation of Street Scenes. In: Daniilidis, K., Maragos, P., Paragios, N. (eds.) ECCV 2010, Part V. LNCS, vol. 6315, pp. 561–574. Springer, Heidelberg (2010)
12. Zhang, H., Fang, T., Chen, X., Zhao, Q., Quan, L.: Partial similarity based nonparametric scene parsing in certain environment. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 2241–2248. IEEE Press, Piscataway (2011)
13. Oliva, A., Torralba, A.: Building the gist of a scene: the role of global image features in recognition. Prog. Brain Res. 155, 23–36 (2006)
14. Bai, X., Sapiro, G.: A geodesic framework for fast interactive image and video segmentation and matting. In: IEEE International Conference on Computer Vision, pp. 1–8 (2007)
15. Price, B., Morse, B., Cohen, S.: Geodesic Graph Cut for Interactive Image Segmentation. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 3161–3168. IEEE Press, Piscataway (2010)
16. Gulshany, V., Rotherz, C., Criminisiz, A., Blakez, A., Zisserman, A.: Geodesic Star Convexity for Interactive Image Segmentation. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 3129–3136. IEEE Press, Piscataway (2010)

17. Chen, X., Zhao, D., Zhao, Y., Lin, L.: Accurate semantic image labeling by fast geodesic propagation. In: IEEE International Conference on Image processing, pp. 4021–4024. IEEE Press, Piscataway (2009)
18. Oliva, A., Torralba, A.: Modeling the shape of the scene: a holistic representation of the spatial envelope. Int. J. Comput. Vis. 42, 145–175 (2001)
19. Xiao, J., Quan, L.: Multiple view semantic segmentation for street view images. In: Proceedings of 12th IEEE International Conference on Computer Vision, pp. 686–693. IEEE Press, Piscataway (2009)
20. Xiao, J., Fang, T., Zhao, P., Lhuillier, M., Quan, L.: Image-based street-side city modeling. ACM Trans. Graph. 28, 1–12 (2009)
21. Arbelaez, P., Maire, M., Fowlkes, C., Malik, J.: Contour Detection and Hierarchical Image Segmentation. IEEE Trans. Pattern Anal. Mach. Intell. 33, 898–916 (2011)
22. Shotton, J., Winn, J.M., Rother, C., Criminisi, A.: *TextonBoost*: Joint Appearance, Shape and Context Modeling for Multi-class Object Recognition and Segmentation. In: Leonardis, A., Bischof, H., Pinz, A. (eds.) ECCV 2006, Part I. LNCS, vol. 3951, pp. 1–15. Springer, Heidelberg (2006)
23. Breiman, L.: Random forests. Mach. Learn. 45, 5–32 (2001)
24. Liaw, A., Wiener, M.: Classification and Regression by randomForest. R News 2, 18–22 (2002)
25. Jaiantilal, A.: Classification and regression by randomforest-matlab (2009), `http://code.google.com/p/randomforest-matlab`
26. Yatziv, L., Bartesaghi, A., Sapiro, G.: O(n) implementation of the fast marching algorithm. J. Comput. Phys. 212, 393–399 (2006)
27. Brostow, G., Fauqueur, J., Cipolla, R.: Semantic Object Classes in Video: A High-Definition Ground Truth Database. Pattern Recognit. Lett. 30, 88–97 (2009)
28. Bileschi, S.: CBCL streetscenes challenge framework (2007), `http://cbcl.mit.edu/software-datasets/streetscenes/`