Detecting Changes in Images of Street Scenes

Jana Košecka

George Mason University, Fairfax, VA, USA

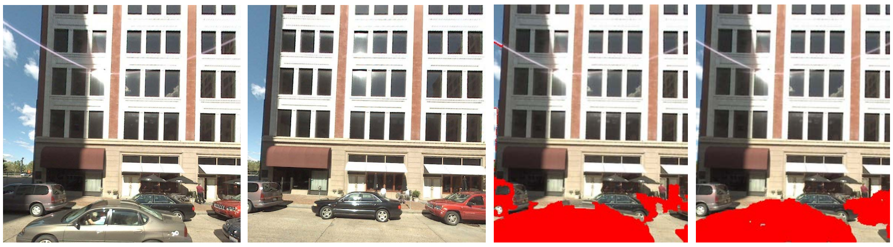
Abstract. In this paper we propose an novel algorithm for detecting changes in street scenes when the vehicle revisits sections of the street at different times. The proposed algorithm detects structural geometric changes, changes due to dynamically moving objects and as well as changes in the street appearance (e.g. posters put up) between two traversal times. We exploit geometric, appearance and semantic information to determine which areas have changed and formulate the problem as an optimal image labeling problem in the Markov Random Field framework. The approach is evaluated on street sequences from 3 different locations which were visited multiple times by the vehicle. The proposed method is applicable to monitoring and updating models and images of urban environments.

1 Introduction

Services like Google StreetView and GoogleEarth are becoming great resource for navigation and search of the constantly growing number of street locations. From the research standpoint these large image (video) datasets continue to pose novel computer vision challenges. In the context of this domain several techniques have been developed for vision based pose estimation, localization and loop closure detection using stereo, monocular or omnidirectional views. Development of robust solutions to these problems tackled the challenges related to the large scale of these datasets and as well as difficulty of lighting conditions due to often low resolution of images and uncontrolled image acquisition environments. The existing solutions exploited the advancements in structure and motion estimation techniques, dense multi-view 3D reconstruction and wide baseline matching and efficient indexing for large scale location recognition. Examples of these can be found in [1], [2], [3], [4], [5], [6], [7] and references therein.

With the success of these services maintenance of 3D city models and associated image panoramas is of importance. At the scale of the city many structural geometric changes (e.g. structures are raised and put down) and appearance changes (e.g. new posters are raised or facades of the buildings modified) happen over larger periods of time. Due to the scale of these datasets the development of automated methods for updating such models or monitoring and reporting the change is of importance. This work focuses on detecting changes in street scenes from images acquired by a moving vehicle. To quantify the amount of change at the level of images we formulate the change detection as optimal labeling problem in Markov Random Field framework, where regions of newly acquired images are labelled into two categories: changed or unchanged.

K.M. Lee et al. (Eds.): ACCV 2012, Part IV, LNCS 7727, pp. 590–601, 2013.



(a) (b) result and ground truth

Fig. 1. Left: (a) and (b) are images of a single location visited at different times; changes are present due to moved cars. Right: is the change detection results obtained by our method and the ground truth. We crop the boundaries for the visualization of the results to visualize only parts which are common in both views.

Contribution. The proposed algorithm for change detection in Street View™ images exploits geometric, appearance and semantic information to determine which areas in the image have changed. In the first stage of our approach we recover a coarse 3D geometry of the scene and register the novel views with the previously acquired reference images of the location. The coarse geometric registration is followed by an appearance transfer stage, where the image regions of a novel view are reprojected to the closest view captured at previous time and their appearance consistency is quantified. In the last stage we exploit semantic content of both previous and current views to gather additional evidence about the change hypotheses. These sources of evidence and integrated in the final energy minimization framework. Depending on whether the changes are structural (building went down), appearance (billboards) or just temporary presence of dynamically moving objects (pedestrians, cars) additional processing steps can be invoked to update 3D geometric models, or Street View images. The example results of the proposed approach can be found in Figure 1.

2 Related Work

The problem of mapping and maintaing models of environments is of fundamental importance for continuous operation in urban environments. Depending on the application domain various instances of this problem have been considered in the autonomous robot localization and mapping communities and surveillance communities. In the surveillance setting the change detection problem is often formulated as 2D-2D image comparison and typically assumes static cameras focusing on the problem of background subtraction [8]. Review of different approaches can be found in 2D images [9]. The methods based on purely 2D information have been found sensitive of changes in illumination and weather conditions. In the work of [10] authors proposed to learn a probabilistic appearance model for a 3D scene and formulated the change detection problem in 3D using voxel based representation of the world. The proposed per voxel appearance model was an extension of mixture of Gaussians estimated from reprojected pixel intensities. In more recent work of [11] authors focus on geometric changes only. They assume the availability of an accurate 3D model of the scene and use the images and their reprojections to new views to generate hypotheses about consistency of the new images with the 3D model. The final inference was formulated in the Markov Random Field framework, where the graph was induced by a 3D voxel grid and the evidence about the voxel change was computed by counting inconsistently projected regions.

Earlier works in the robotics community considered issues of dynamically changing environments in the context of simultaneous localization and mapping problem. These methods typically rely purely on 3D geometry or 2D occupancy maps. [12] addressed the problem of localization in dynamic environments in an on-line manner using occupancy grid based representation, where both static and dynamic parts of the environment were represented in terms of separate occupancy grids. In the work of [13] the issue of dynamic changes have been tackled at the level of entire map using map differencing techniques and Expectation Maximization Algorithm; [14] proposed a method for on-line detection and identification of moving objects assuming ideal localization. The proposed work is the closest to [10,11] approaches to change detection. We also exploit information about 3D geometry and relative poses between the views, but formulate the final inference problem in 2D space of the new image instead of 3D voxel grid. In addition to geometric geometric changes, we consider capturing changes in environment appearance, such as posters or billboards put up or removed.

Instead of considering freely moving camera, we tackle the change detection problem using Street View image panoramas acquired by moving vehicle. The problem of change detection in this context is relevant for navigation and loop closing, where areas of the city are revisited by the vehicle. These omnidirectional views make the problem of image registration better conditioned despite their lower resolution, but also pose some challenges due to dramatic appearance variations and presence of large repetitive structures. The change detection algorithms are applied only to the side views of the panorama, oriented 90^o from heading direction of the vehicle.

Outline. In Section 3 we discuss the techniques for pose estimation used to register the views of a location acquired at different times. Section 4 describes our algorithm for change detection, detailing the geometric, appearance and semantic cues. We formulate the problem as optimal image labeling in Markov Random Field framework, followed by the results and conclusions in Section 5.

3 Preliminaries

The Street View images have been acquired by standard perspective cameras aligned in a circle. Our panorama is composed of four perspective images covering 360^o horizontally and 127^o vertically. We have multiple frames of each location available. Examples of images from 3 different locations at different times and changes we consider are in Figure 2.

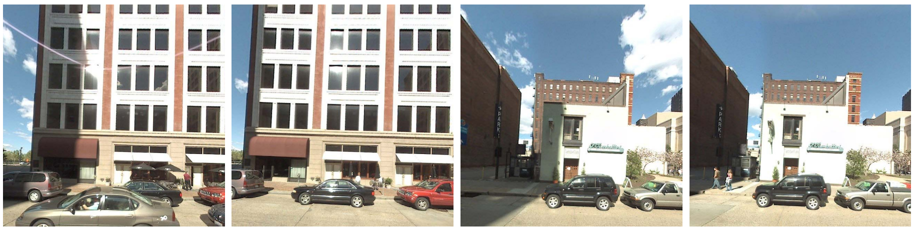


Fig. 2. Images of example locations and the same locations revisited at different time of the day

Given a reference sequences of images $I^r_i, ..., I^r_j$ and a sequence acquired at later time $I^q_k, ..., I^q_l$, the first stage of our algorithm recovers the relative pose between the views in the reference sequence and recovered 3D structure. We employ standard visual odometry pipeline to recover relative poses and one single global scale of these views from the images. We use the wide baseline matching using SIFT features between each consecutive image pair along the sequence. The prismatic representation of the omnidirectional image allows us to construct corresponding 3D rays $\mathbf{p}, \mathbf{p}'$ for established tentative point matches $\mathbf{x}^q_t \leftrightarrow \mathbf{x}^q_{t+1}$. The tentative matches are validated through RANSAC-based epipolar geometry estimation formulated on their 3D rays, $\mathbf{p}'^\top \mathtt{E}\, \mathbf{p} = 0$, yielding thus the essential matrix $\mathtt{E}$ [15]. Improved convergence of RANSAC can be achieved if rays are sampled uniformly from each of four subparts of the panorama. It has been shown in the past that this yields more accurate estimates of pose [16] even in the absence of bundle adjustment. We denote the two consecutive novel views I^q_t and I^q_{t+1} and the nearest reference view I^r_k. We establish correspondences $\mathbf{x}^q \leftrightarrow \mathbf{x}^r$ between the novel view I^q_t and the closest reference view I^r_k and compute the pose from the essential matrix between the views. For solving the scales of translations between consecutive pairs of images and the reference view we set the norm of the translation for the first novel pair to be 1. Scale of the translation is estimated by a linear closed-form 1-point algorithm on corresponding 3D points triangulated from the query image pair and the reference view.

Given the registered set of novel views, we compute a coarse 3D structure of the scene. Instead of employing the full 3D dense reconstruction pipeline, we segment the image into small superpixels and establish correspondences between each centroid of the superpixel and it's consecutive view in the query sequence. Due to the fact that these frames are relatively close in time and the displacements are small, we used dense optical flow method [17] to establish the correspondences and using the median flow of pixels in the superpixel as displacement. 3D position of the superpixel centroid is then triangulated yielding a coarse 3D model. The quality of the model can be substantially improved using more advanced multi-view stereo reconstruction techniques. An example of 3D reconstruction at the superpixel level can be seen in Figure 3.

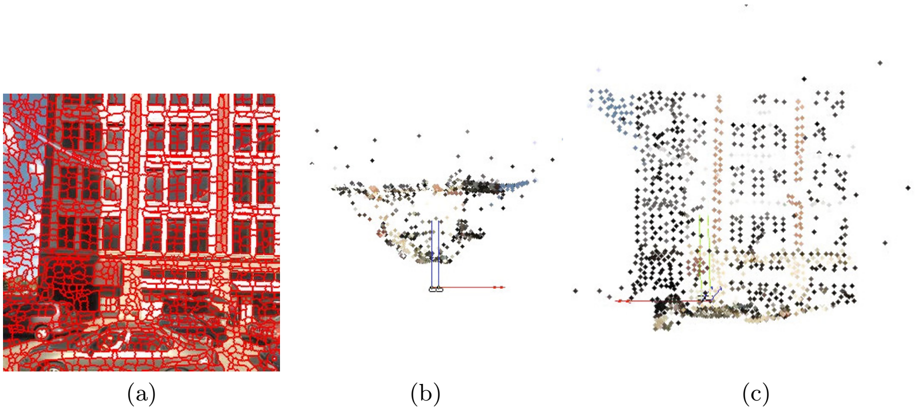


Fig. 3. (a) Image segmented into small watershed superpixels; (b) Bird's eye view of 3D reconstruction of elementary superpixels; (c) Side view of the same 3D structure

4 Change Detection

Previous section discussed the components of our system for the image alignment and a coarse 3D reconstruction. We propose to formulate the change detection problem as an optimal labeling problem in MRF framework, where we will seek an optimal label assignment 0 or 1 to each superpixel signifying whether the region changed (1) or remained the same (0). We seek to maximize the posterior probability of the labels $\mathbf{L} = \{0, 1\}$ given image observations. The label likelihoods and joint prior are expressed as unary and binary functions used in a second-order MRF framework. This maximization problem is equivalent to the energy minimization re-written in a log-space and has the following form

$$\operatorname*{argmin}_{\mathbf{L}} \Big(\sum_{s_i \in S} \phi^U(s_i) + \lambda_s \sum_{(i,j) \in \mathcal{G}} \phi^P(s_i, s_j) \Big). \tag{1}$$

where the terms $\phi^U(s_i)$ are unary potentials quantifying the amount of change in a superpixel and $\phi^P(s_i, s_j)$ measure the pairwise consistency between the neighboring superpixels. The structure of MRF is induced by image superpixels s_i. These in our case are computed by watershed segmentation on Laplacian of Gaussian (LoG) interest points as seeds and can be seen in Figure 3(a). LoG interest points are selected as extrema of 4 level Laplacian of Gaussian pyramid described in more details [18]. This method of seed selection places interest points densely yielding small regions when followed by watershed segmentation. These elementary regions typically do not straddle boundaries between different classes and naturally contain semantically meaningful object or scene primitives. Furthermore, they dramatically reduce computational complexity of 3D reconstruction and an MRF inference. We describe the form of unary and binary potentials next.

4.1 Unary Term

Geometry and Appearance. One component of the unary term quantifies the geometric and appearance change for each superpixel. To capture the appearance of superpixel s_i^q in the query view I_t^q, each superpixel is characterized by SIFT descriptor d_i computed at the superpixel's center. We use the 3D reconstruction of the superpixel s_i^q and the pose between the novel view and the closest reference view to find the corresponding superpixel in the reference view s_j^r and its associated descriptor d_j. As a measure of similarity $dist(s_i, s_j)$ we use the cosine of the angle between the descriptor of the query superpixel s_i and the superpixel s_j which is nearest to the location of the reprojected centroid of s_i in the reference view I^r.

$$\phi_{SIFT}(s_i) = \begin{cases} \exp\left(-\frac{(1-d(s_i,s_j))^2}{2\sigma^2}\right), & \text{if } r_{err}(s_i) < \tau \\ 0.5 & \text{otherwise} \end{cases} \tag{2}$$

where $r_{err}(s_i)$ is the reprojection error of the 3D reconstruction of i^{th} superpixel, from the two consecutive views of the novel sequence. In our experiments we use $\sigma = 0.25$ and $\tau = 1$ pixels. This strategy for appearance transfer is similar to the methods used for semantic labeling explored by SIFT flow [19], but the process of finding correspondences is eased by the availability of a coarse 3D geometry. Figure 5c shows an example visualizing different confidence values of appearance changes. Note that darker areas of lower confidence are due to either dramatic lighting changes or large reprojection errors errors caused by dynamically moving objects (e.g. cars).

Semantic Labeling. In order to gather additional evidence to support the final inference process, we propose to incorporate evidence about different semantic labels associated with image regions. In the next section we describe our approach to semantic labeling and describe how to incorporate the evidence about semantic labels into the final inference stage. Various approaches to semantic labeling with the focus on street scenes include works of [20], [21], [22] and [23]. In the context of our domain we consider the problem of assigning semantic labels *ground, sky, building, car, tree* to different regions of the image. We choose the superpixels obtained by color based over segmentation scheme proposed in [24].

The choice of features has been adopted from [25] where each superpixel is characterized by location and shape (position of the centroid, relative position, number of pixels and area in the image), color (color histograms of RGB, HSV values and saturation value), texture (mean absolute response of the filter bank of 15 filters and histogram of maximum responses) and perspective cues computed from long linear segments and lines aligned with different vanishing points. The entire feature vector is of 194 dimensions. In order to compute the likelihood of individual superpixels, we use boosting [26]. In our implementation, each strong boosting classifier has 15 decision trees and each of the decision trees has 6 nodes. The classifier was trained using randomly selected half of the 320 side view dataset similar to [27] and [28]. The other half of the dataset is used for

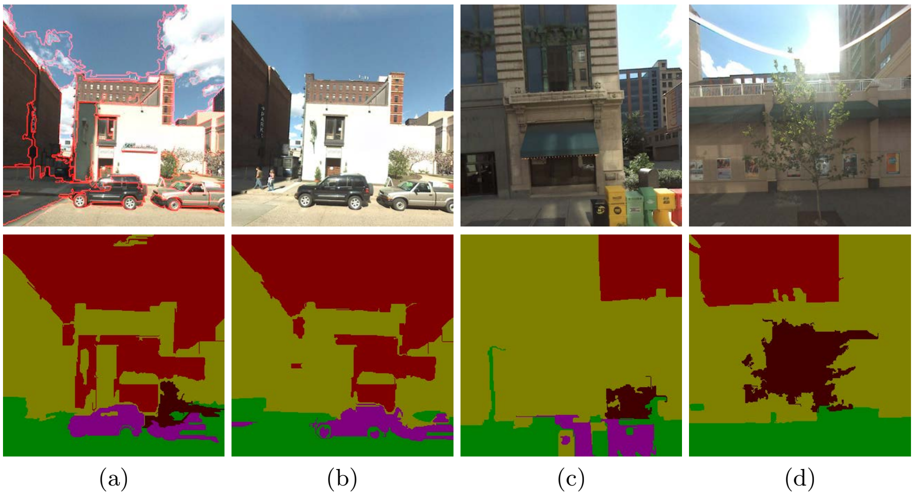


Fig. 4. Top row: (a) Example of the color-based over segmentation using method of [24] superpixel boundaries are marked by red color; (b)-(d) example street views. Bottom row: Semantic labeling result for the given over segmentation and boosting classifier, only data term is visualized. Note that due to the crude initial segmentation, several image regions are misclassified. (e.g. shaded are of the building in (a) is misclassified as sky (due to the same color). (c) mailboxes are classified as car. The color coding is the following: building: yellow, car: purple, ground: green, sky: red, tree: brown.

Table 1. Category wise accuracy of boosting classifier; global and average accuracy in % correct

System	**build.**	**car**	**ground**	**sky**	**tree**	**glob.**	**aver.**
[28]	89.1	56.4	89.6	97.1	69.7	88.4	80.4
[27]	95.3	40.5	96	92.5	41.4	93.2	73.1
our	96.4	68.3	94.4	97.2	48.9	94.4	81

testing. Each pixel of an image was assigned one of the five classes or *void* if it does not fall into any of the categories. Although the semantic labeling is not the final goal of this work, we have compared the performance of the boosting classifier and with the state of the art systems in Supervised Label Transfer [28] and Non-parametric scene parsing [27] in Table 1. Note that despite the fact that we do not use any MRF regularization stage, our approach outperforms the previously proposed methods for the categories of interest. Some examples of the results of semantic segmentation are in Figure 4.

While for the chosen categories the approach performs quite well due to rich features and large regions of support, there are still many cases where the label assignments are incorrect, see Figure 4 or 5. One source of errors is the local ambiguity of the region as described by the features and another is the errors of initial over segmentation into superpixels.

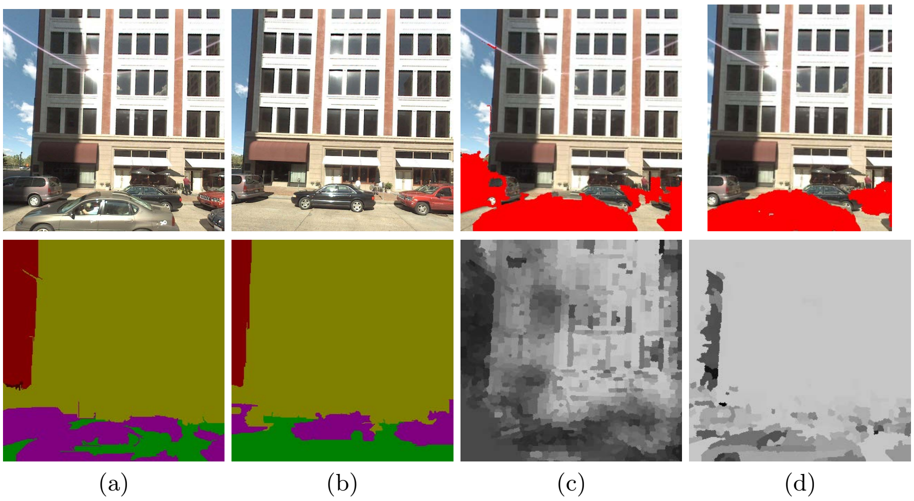


Fig. 5. Change detection example and ingredients. Top row: query view, reference view, result and ground truth information. Bottom row: (a), (b) semantic labels of the two views, (c)confidence map: distance between descriptor of a superpixel and its reprojected counterpart in the previous view, (d) KL divergence between semantic layout of each superpixel and its reprojected counterpart.

To quantify the amount of semantic change between two views, we use the entire label distribution obtained for each large superpixel for both the query view and the reference view. The output of the boosting classifier returns confidence values $f^k(s_i)$ for each superpixel belonging to a particular class k, which can be interpreted as probability by passing it through a sigmoid function

$$p_k = P(l = k|f(s_i)) = \frac{1}{1 - \exp(-f(s_i))}.$$

This gives a probability distribution of labels for each superpixel $p_q = [p_1, \dots p_k]$ in the query view and reference view $p_r = [p_1, ...p_k]$. The amount of change can then be related to the difference between the two distributions. Commonly used difference is the Kullback-Leibler Divergence of p_r and p_q defined as

$$\phi_{KL}(s_i) = \frac{1}{k} \sum_{i=1}^{k} p_q(i) \log \frac{p_q(i)}{p_r(i)}.$$

This difference is computed for each registered small superpixel s_i and its reprojected counterpart s_j in the reference view.

The final form of the unary term then becomes weighted combination of the semantic and the appearance information

$$\phi^U(s_i) = \alpha\phi_{SIFT}(s_i) + (1 - \alpha)\phi_{KL}(s_i) \ . \tag{3}$$

In our experiments the we find the optimal α by validation with respect to the ground truth data as $\alpha = 0.7$. We have a small dataset of 10 ground truth views, from 3 different locations, where we manually annotated the regions in the novel query views, which do not appear in the closest reference view. Ideally this term should be determined in a data driven way as the confidence in semantic segmentation can vary dramatically for different query views.

4.2 Pairwise Term

In our case we choose simple data driven prior based on color differences. The joint prior or the smoothness term, is approximated by pairwise potentials as

$$\phi_{smooth}(s_i, s_j) = \exp\Big(\sum_{(i,j)\in\mathcal{E}} g(i,j)\Big), \tag{4}$$

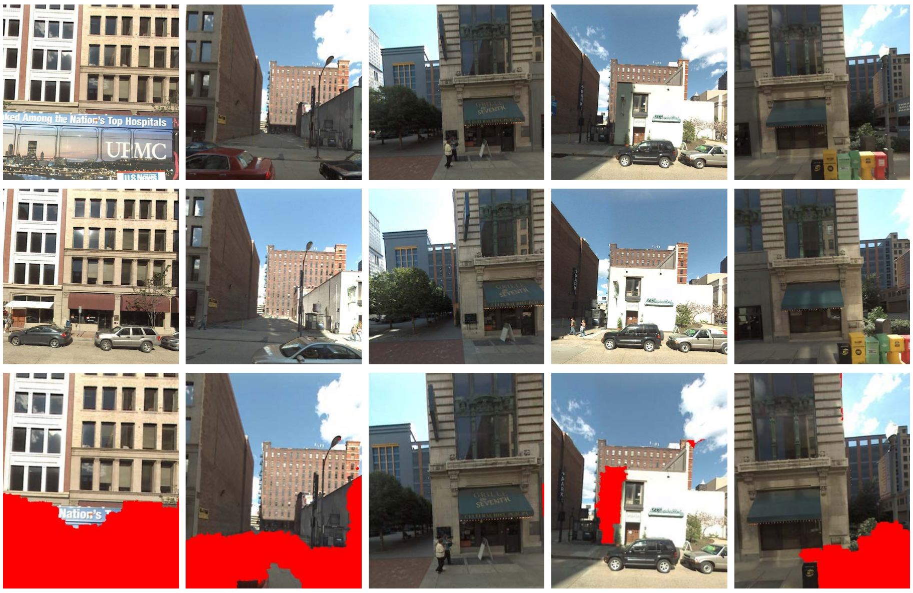
where the pairwise affinity function g is defined as

$$g(i,j) = \begin{cases} 1 - e, & \text{iff } l_i = l_j \\ \delta + e, & \text{otherwise,} \end{cases} \tag{5}$$

with $e = \exp(-\|\mathbf{c}_i - \mathbf{c}_j\|^2/2\sigma^2)$, where $\mathbf{c}_i$ and $\mathbf{c}_j$ are 3-element vectors of mean colors expressed in the Lab color space for i-th and j-th superpixel, respectively, and σ is a parameter set to 0.1. The set $\mathcal{E}$ contains all neighboring superpixel pairs. The smoothness term is a combination of the Potts model penalizing different pairwise labels by the parameter δ and a color similarity based term. The aim is on one side to keep the same labels for neighboring superpixels, and on the other, to penalize same labels if they have different color. The scalars λ_s and δ weigh the importance of the terms (set to 1 and 0.2 in our experiments).

We perform the inference in the MRF by efficient and fast publicly available MAX-SUM solver [29] based on linear programming relaxation and its Lagrangian dual. Figure 6 shows some examples of the proposed change detection algorithm. We achieved 73.5% average accuracy of the change detection, averaged over 3 different locations.

There are two sources of inaccuracies in our method. As mentioned at the beginning we rely on a coarse 3D reconstruction, where correspondences are established using optical flow techniques. While the small baseline makes the problem of establishing correspondences easier there are still errors in the areas of uniform intensities and occlusions. These errors are further propagated to the reconstruction stage. Due to the fact that we use simple linear triangulation without additional regularization stage, 3D coordinates of superpixels have errors. These errors are propagated to novel views causing incorrect confidences in the appearance change. Some of these issues can be tackled by more robust motion estimation methods which explicitly model occlusion phenomena [30] or more advanced stereo reconstruction techniques. Availability of accurate 3D model would improve the accuracy of the reprojection stage [6]. Note also that we do not explicitly handle dynamically moving objects in the query view pair. In case



(a) (b) (c) (d) (e)

Fig. 6. Examples results of the change detection. The top row are the new query views; middle row are the closest views from the reference databased taken at earlier period of time and bottom row are the results of the change detection algorithm. Columns (d) and (e) show mistakes of the algorithm, which are due to differences in semantic labeling shown in 4.

the extent of moving objects and their motion is small their effect on pose estimation and 3D reconstruction is negligent. Additional challenge comes from the fidelity of the semantic segmentation. While the proposed method is comparable with the state of the art methods, it often produces incorrect labels. These unreliable label distributions are further propagated to the final optimization stage. More advanced methods for semantic segmentation would further improve the estimated label confidences.

5 Conclusions

We have presented a novel algorithm for change detection which combines geometric, appearance and semantic information. Street View images are acquired by a moving vehicle and densely sampled making the viewpoint changes between the new and old views constrained. This makes the use of patch based descriptors and their invariance properties feasible. In order to tackle the difficult appearance variations due to illumination changes, reflections and inter-reflections we use the hypotheses generated by semantic segmentation algorithm. This algorithm uses over-segmentation in to larger superpixels and exploits statistics (features) computed over larger spatial regions. In the current approach the evidence is integrated in a single global MRF inference. Further improvements can be achieved by using more advanced 3D reconstruction methods as well better semantic segmentation strategies which exploit geometry and temporal continuity.

Acknowledgement. The author would like to thank the anonymous reviewers for their valuable comments. This project has been supported by Army Research Office Grant W911NF-1110476 and NGA Initiation Grant.

References

1. Anati, R., Daniilidis, K.: Constructing topological maps using Markov Random Fields and Loop Closure Detection. In: NIPS, pp. 37–45 (2009)
2. Cummins, M., Newman, P.: Highly scalable appearance-only slam - FAB-MAP 2.0. In: Robotics Science and Systems, RSS, Seattle, USA (2009)
3. Kumar, A., Tardif, J.P., Anati, R., Daniilidis, K.: Experiments on visual loop closing using vocabulary trees. In: CVPR Workshop, pp. 1–8 (2008)
4. Jones, E., Soatto, S.: Visual-inertial navigation, mapping and localization: A scalable real-time causal approach. International Journal of Robotics Research (2011)
5. Micusik, B., Košecka, J.: Piecewise planar city modeling from street view panoramic sequences. In: IEEE Conference on Computer Vision and Pattern Recognition (2009)
6. Pollefeys, M., Nister, D., Frahm, J.M.: Detailed realtime urban 3D reconstruction from video. Int. Journal on Computer Vision (2008)
7. Schindler, G., Brown, M., Szeliski, R.: City-scale location recognition. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 1–7 (2007)
8. Stauffer, C., Grimson, E.: Adaptive background mixture models for real-time tracking. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 246–252 (1999)
9. Radke, R., Andra, S., Al-Kofani, O., Roysam, B.: Image change detection algorithms. IEEE Transactions of Image Processing (2005)
10. Pollard, T., Mundy, J.: Change detection in 3D world. In: IEEE Conference on Computer Vision and Pattern Recognition (2011)
11. Taneja, A., Ballan, L., Pollefeys, M.: Image based detection of geometric changes in urban environments. In: IEEE Conference on Computer Vision and Pattern Recognition (2011)
12. Wolf, D., Sukhatme, G.: Online simultaneous localization and mapping in dynamic environments. In: IEEE Conference on Robotics and Automation (2004)
13. Hahnel, D., Triebel, R., Burghard, W., Thrun, S.: Map building with mobile robots in dynamic environments. In: IEEE Conference on Robotics and Automation (2002)
14. Biswas, R., Limetkai, B., Sanner, B., Thrun, S.: Towards object mapping in nonstationary environments with mobile robots. In: International Conference on Intelligent Robots and Systems (2002)
15. Ma, Y., Soatto, S., Košecka, J., Sastry, S.: Invitation to 3D vision: From Images to Geometric Models. Springer (2002)
16. Tardif, J.P., Pavlidis, Y., Daniilidis, K.: Monocular visual odometry in urban environments using an omnidirectional camera. In: Proc. of IEEE Int. Conf. on Intelligent Robots and Systems, IROS (2008)

17. Brox, T., Malik, J.: Large displacement optical flow: descriptor matching in variation motion estimation. IEEE Transactions on Pattern Analysis and Machine Intelligence 33, 500–513 (2011)
18. Lowe, D.: Distinctive image features from scale-invariant keypoints. Int. Journal on Computer Vision 60, 91–110 (2004)
19. Liu, C., Yuen, J., Torralba, A., Sivic, J., Freeman, W.T.: SIFT Flow: Dense Correspondence across Different Scenes. In: Forsyth, D., Torr, P., Zisserman, A. (eds.) ECCV 2008, Part III. LNCS, vol. 5304, pp. 28–42. Springer, Heidelberg (2008)
20. Tighe, J., Lazebnik, S.: Superparsing: Scalable nonparametriv image parsing with superpixels. In: Proc. of European Computer Vision Conference (2010)
21. Huang, Q., Han, M., Wu, B., Ioffe, S.: A hierarchical conditional random field model for labeling and segmenting images of street scenes. In: IEEE Conference on Computer Vision and Pattern Recognition (2011)
22. Ladicky, L., Sturgess, P., Russell, C., Sengupta, S., Bastanlar, Y., Clocksin, W., Torr, P.H.: Joint optimisation for object class segmentation and dense stereo reconstruction. In: Proc. of British Machine Vision Conference (2010)
23. Xiao, J., Quan, L.: Multiple view semantic segmentation for street view images. In: Proc. of Int. Conference on Computer Vision, pp. 1–8 (2009)
24. Felzenszwalb, P., Huttenlocher, D.: Efficient graph-based image segmentation. Int. Journal on Computer Vision 59, 167–181 (2004)
25. Hoeim, D., Efros, A., Hebert, M.: Recovering surface layout from an image. Int. Journal on Computer Vision, 151–172 (2007)
26. Schapire, R.E., Singer, Y.: Improved boosting using confidence-rated predictions. Machine Learning 37, 297–336 (1999)
27. 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 (2011)
28. Zhang, H., Xiao, J., Quan, L.: Supervised label transfer for semantic segmentation of street scenes. In: Proc. of European Computer Vision Conference, pp. 561–574 (2010)
29. Werner, T.: A linear programming approach to Max-sum problem: A review. PAMI 29, 1165–1179 (2007)
30. Ayvaci, A., Raptis, M., Soatto, S.: Sparse occlusion detection with optical flow. International Journal of Computer Vision 97 (2012)