Geometry-Based Superpixel Segmentation
Introduction of Planar Hypothesis for Superpixel Construction

Keywords: Image Segmentation, Superpixel, Planar hypothesis.

Abstract: Superpixel segmentation is widely used in the preprocessing step of many applications. Most of existing methods are based on a photometric criterion combined to the position of the pixels. In the same way as the SLIC method, based on k-means segmentation, a new algorithm is introduced. The main contribution lies on the definition of a new distance for the construction of the superpixels. This distance takes into account both the surface normals and a similarity measure between pixels that are located on the same planar surface. We show that our approach improves over-segmentation, like SLIC, i.e. the proposed method is able to segment properly planar surfaces.

1 INTRODUCTION

The image segmentation problem consists in partitioning an image into homogeneous regions supported by groups of pixels. This approach is commonly used for image scene understanding (Mori, 2005; Gould et al., 2009). Obtaining a meaningful semantic segmentation of a complex scene containing many objects: rigid or deformable, static or moving, bright or in a shadow is a challenging problem for many computer vision applications such as autonomous driving, traffic safety or mobile mapping systems.

Over the past decade, superpixels have been widely used in order to provide coherent and reliable over-segmentation, i.e. each region contains only a part of the same object and respects the edges of this object (Felzenszwalb and Huttenlocher, 2004; Achanta et al., 2012). Superpixels are intermediate features providing spatial support that brings more information than just using pixels. Superpixels decomposition also allows to reduce problem complexity (Arbelaez et al., 2009). Consequently, it is a useful tool to understand and interpret scenes. Existing superpixels approaches take into account a photometric criterion, color differences between pixels have to be minimal in the same superpixel, and a shape constraint that is based on the space distance between pixels. Approaches based only on these two criteria can provide superpixels that cover two surfaces with different orientations. On figure 1, there is a such superpixel on the edge of the cube, corresponds to a non-planar area. It is difficult to semantically classify a superpixel that represents two different 3D entities.

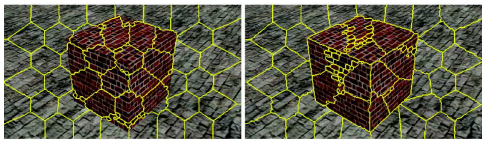

Figure 1: Superpixels comparison between k-means approach (left) with a hard compactness fixed at $m = 40$ and the proposed approach (right) with $m = 5$.

In order to take into account this kind of difficulties, in single view segmentation methods, geometric criteria are introduced such as the horizon line or vanishing points (Hoiem et al., 2005; Saxena et al., 2008; Gould et al., 2009). Even if some geometry information is introduced, these existing approaches do not integrate it in the over-segmentation process but only as a post-processing step to classify superpixels. It means that errors on superpixels, i.e. superpixels that contain multiple surfaces with different orientations might be propagated and not corrected.

In the case of calibrated multi-view images, redundant information are available. Consequently, the geometry of the scene can be exploited to strengthen the scene understanding. For example, in man-made environment, it is common to make a piece-wise pla-

nar assumption to guide the 3D reconstruction (Bartoli, 2007; Gallup et al., 2010). This kind of information is combined with superpixels in (Mičušík and Košecká, 2010) but, in fact, the geometric information is not integrated in the construction of the intermediate entities (superpixel or face mesh) and errors of this over-segmentation are also propagated.

In this article, we focus on the multi-view images context. In order to obtain superpixels that are coherent with the scene geometry, we propose to integrate a geometric criteria in superpixels construction. The proposed algorithm follows the same steps as the well known SLIC, *Simple Linear Iterative Clustering* approach (Achanta et al., 2012) but the aggregation step takes into account the surface orientations and the similarity between two consecutive images. In §2, we present a brief state of the art on superpixels constructors. Then, an overview of the proposed framework is presented, followed by the extraction of geometric information and its integration in a k-means superpixels constructor. Finally, experiments on synthetic data are presented.

2 SUPERPIXELS

In the context of superpixels construction, we propose to distinguish three kinds of methods: graph-based approaches (Felzenszwalb and Huttenlocher, 2004; Moore et al., 2008), seed growing methods (Levinshtein et al., 2009) and methods based on k-means (Comaniciu and Meer, 2002; Achanta et al., 2012). We will focus on the last set of methods and in particular on (Achanta et al., 2012) because this method provides in three simple steps presented in the following paragraph, uniform size and compact superpixels, widely used in the literature (Wang et al., 2011). After briefly describing this method, we analyze its advantages and drawbacks. This allows us to highlight the significance of the compactness criterion put forward in (Schick et al., 2012).

K-means Superpixel – SLIC (Achanta et al., 2012) is a single color image over-segmentation algorithm based on k-means superpixels. It provides uniform size superpixels, that means they contain approximately the same number of pixels. SLIC is based on a 5 dimensional k-means clustering, 3 dimensions for the color in the Lab color space and 2 for the spatial features x, y corresponding to the position in pixel. The algorithm follows these three steps:

1. Seeds initialization on a regular grid of $S \times S$ and distributed on 3×3 neighborhood to reach the lower local gradient;
2. Compute iteratively superpixels on a local window until convergence:
   (a) Aggregate pixels to a seed by minimizing D_{SLIC} distance (1) on a searching window of size $2S \times 2S$;
   (b) Update position of cluster centers by calculating the mean on each superpixel;
3. Enforce connectivity by connecting small entities using connected component method.

Two parameters need to be set for SLIC, the approximate desired number of superpixels K, as well as in most of the over-segmentation method, the weight of the relative importance between spatial proximity and color similarity m which is directly linked to the compactness criterion, see equation (1).

Energy Minimisation – The energy-distance to minimize between a seed and a pixel that belongs to the window centered on the seed is defined by:

$$D_{SLIC} = \sqrt{d_c^2 + \frac{m^2}{S^2} d_s^2} \tag{1}$$

where

- d_c and d_s are color and spatial distance,
- m is the compactness weight,
- $S = \sqrt{\frac{N}{K}}$
- N is the number of pixels in the image,
- K is the number of superpixels asked.

In the case of a color picture, the distance are defined as following:

$$\begin{aligned} d_c(p_j, p_i) &= \sqrt{(l_j - l_i)^2 + (a_j - a_i)^2 + (b_j - b_i)^2} \\ d_s(p_j, p_i) &= \sqrt{(x_j - x_i)^2 + (y_j - y_i)^2}. \end{aligned} \tag{2}$$

Analysis – The superpixel compactness and connectivity are two properties that are desirable for superpixels. On one hand, the compactness (Levinshtein et al., 2009; Moore et al., 2008; Achanta et al., 2012; Schick et al., 2012) of a superpixel can be defined by the quotient between the area and the perimeter of the shape. In figure 2 shows the influences of the weight on the space distance d_s, in SLIC and how it impacts the compactness. Moreover, the k-means superpixel algorithm enforces to use pixels in a local window. It sets the upper value of the compactness to the size of the searching window. On the other hand, a superpixel is connected if all its pixels belong to a unique connected entity. This property is enforced in the third step of the algorithm

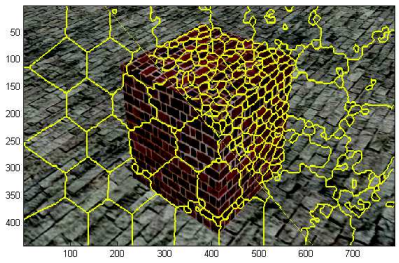

Figure 2: K-means superpixels compactness comparison with a small number (50) of desirable superpixel : bottom-left hard compactness at $m = 40$ and top-right a soft compactness at $m = 5$. For the hard compactness, the desirable number of superpixels is almost respected, since for the soft compactness this number is blowed up.

and is relative to the image resolution.

Since we have remarked that existing superpixels methods are usually based on photometric criterion with some topology property in the image space, in the next part, we propose a variant of k-means superpixels constructor on two images. This is done by integrating the geometric information in order to obtain superpixels coherent with the scene geometry, conpact even with a small number of representative entities.

3 GEOMETRY-BASED SUPERPIXEL CONSTRUCTION

In this work, we deal with (at least) two images of an urban scene i.e., a scene that is basically piecewise planar. Similarly to related works (Bartoli, 2007), we assume to have at our disposal a sparse 3D reconstruction of the scene, provided by some structure-from-motion algorithm (Wu, 2011). We aim at segmenting the images into superpixels using a method relying on a k-means approach, namely SLIC. Our idea is to integrate in the proposed superpixel constructor the available geometric information.

In this section, we first present the available input data and describe which information can be extracted in order to be exploited in the superpixel constructor. More precisely, we propose to use two maps of the same size than the input images: for each pixel p, the first, called similarity map, measures the planarity of the surface supporting the 3D point that projects into p while the second, called normal map, estimates the normal of this surface. We also explain how these two maps are used as quantitative values to modify the SLIC distance.

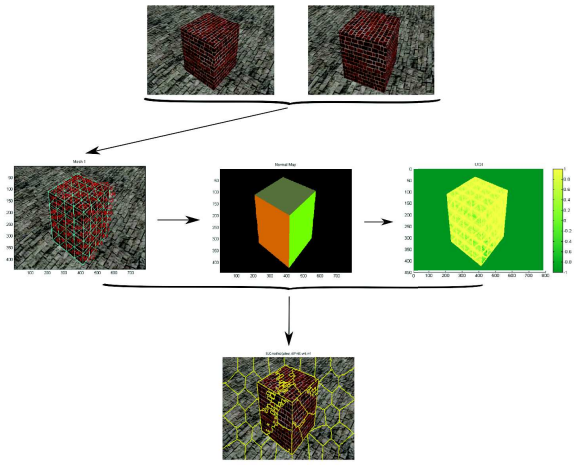

Figure 3: Framework of our proposed over-segmentation method using scene geometry. At the top, the two images I and I'. In the second row: the Delaunay triangulation from 2D interest points matched with the other view; the normal map estimated on the faces of the mesh, helped by the epipolar geometry and the similarity map between both views. The over-segmentation results should be coherent with the scene geometry.

3.1 Input Data

We use two colors and calibrated images I and I'. We denote $\mathtt{P}_I = \mathtt{K}[\mathtt{I}|\mathbf{0}]$ the projection matrix of the reference image I, where $\mathtt{K}$ is the matrix of the intrinsic parameters and $\mathtt{P}_{I'} = \mathtt{K}[\mathtt{R}|\mathbf{t}]$ the projection matrix associated to the image I' where $\mathtt{R}$ is the rotation matrix and $\mathbf{t}$ the translation vector that determines the relative poses of the cameras. More details about the geometric aspects are provided by (Hartley and Zisserman, 2004). A sparse 3D point cloud can be projected in each images through the projection matrix to obtain a set of 2D matched points. We note z and z' a part of the reference images and of the adjacent image. $\tilde{z}$ corresponds to the warped part of the adjacent image estimated by the homography induced by the plane of support of a triangle defined by three points.

3.2 Geometry Extraction

After a presentation of the available input data, we introduce how we extract geometric information from multi-view images in order to exploit it in a k-mean superpixels constructor.

A given 2D Delaunay triangulation on the set of 2D points of interest in the reference image can be applied to the corresponding 3D points. Doing so, enables to estimate 3D plane on each face of the mesh determined by three 3D points.

Normal Map – The normal map associated to the reference image represents for each pixel p_i the normal orientation $\vec{n}_i$ of the plane represented by the face of the mesh in the image. It is a 3D matrix, containing the normalised normals value along the 3D axis in $[-1,1]$. Some missing pixels do not have evaluated normal, those will be considered with $\varnothing$.

Planarity Map – For each triangle, knowing the plane parameters and the epipolar geometry, we can estimate the homography induced by the plane of support. This homography enables to compute the warped image $\tilde{z}$, aligned to the part of the reference image. Then, the two images z and $\tilde{z}$ can be compared using an a full referenced IQA.

An IQA, also called photo-consistency criterion, measures the similarity or the dissimilarity between two images. Two kinds of measures take a huge place in evaluation process results. Those based on Euclidean distance with the well-known Mean Square Error (MSE) and the cosine angle distance-based such as the Structure SImilarity Measure (SSIM) (Wang et al., 2004).

The work of (Author, 2015) shows that measures based on cosine angle differences are more efficient than Euclidean based-distances for planar/non-planar classification. Illustrated in figure 4, when a high similarity is obtained, the representative part corresponds to a good estimation of the parameters of the planar surface, otherwise it is assimilated to a non-planar surface. Non-planar surfaces are difficult to manage because the dissimilarity between z and $\tilde{z}$ can be induced by many difficulties in the scene, such as occlusions, moving objects, specularities. We used UQI (Z. Wang and Bovik, 2002) a specific case of SSIM to classify, with a simple threshold, pixels that belong to a planar surface with a high similarity and those with a low similarity that do not belong to the planar surface. Same as the normal map, the missing pixels that do not belong to the mesh are considered with $\varnothing$ value.

We have presented the two maps containing the 3D geometric information we have extracted. The normal map gives information on the surface orientation since the similarity map validates or rejects the planar assumption.

3.3 Geometry-Based Superpixels

We propose an new energy to be minimized, defined as following:

$$D_{SP} = \sqrt{d_{c_0} + \alpha.d_{s_0}^{\beta} + d_g} \tag{3}$$

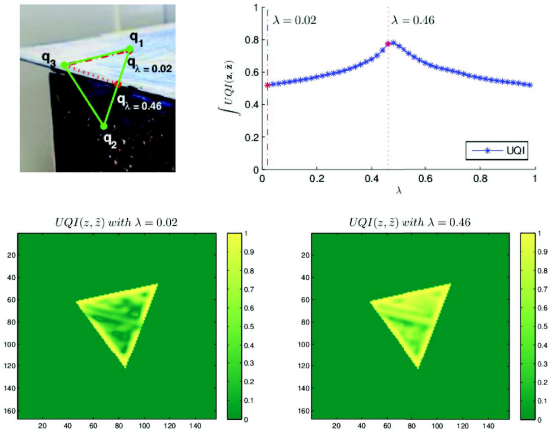

Figure 4: UQI behaviour on a non-planar case. Fist row:the reference image triangle z to which $\tilde{z}$ the warped triangle is compared with the IQU measure. A point q_λ slides from q_1 to q_2 in order to reach good plane parameters. Top-right: Curve of the means similarity value obtained for each λ. Second row: similarity map for two cases. Left: $\lambda = 0.02$ the wrong estimation parameters are used to compute the warped image and a low mean similarity value is obtained. Right: $\lambda = 0.46$ the maximum similarity value is reached with the correct plane parameters estimation where $q_{\lambda=0.46}$ belongs to the two planes intersection.

We add a new term in the distance used to aggregate pixels to a superpixel. This term d_g, takes into account the scene geometry by merging the surface normals orientation map and the similarity map.

$$d_g(p_j,p_i) = d_{\vec{n}}(p_j,p_i).d_{IQA}(p_j). \tag{4}$$

We also define, d_{s_0} and d_{c_0} the normalized distances of d_s and d_c. Let $d_{\vec{n}}$ be the normal distance, measuring the cosine angle between normals in two points. Let d_{IQA} correspond to the positive value of the similarity map. Since dissimilar pixels are rejected cases, we can use a hard threshold, here zero, to remove noise and unmeaning values.

$$\begin{aligned} d_{s_0}(p_j,p_i) &= \frac{d_s}{max(d_s)} \\ d_{c_0}(p_j,p_i) &= \frac{d_c}{max(d_c)} \\ d_{\vec{n}}(p_j,p_i) &= \frac{1+cos(\vec{n}_j,\vec{n}_i)}{2} \\ d_{IQA}(p_j) &= IQA(p_j).\mathbb{1}_{IQA>0}. \end{aligned} \tag{5}$$

The three terms d_{s_0}, d_{c_0} and d_g, of the proposed distance D_{SP} presented equation 3, are illustrated figure 5. The normalisation of d_s and d_c enables to be more aware of the impact of weights α and β on the d_{s_0} term related to the compactness. The curve illustrated in figure 6, shows the influence of these two parameters. α influences this weight between compact-

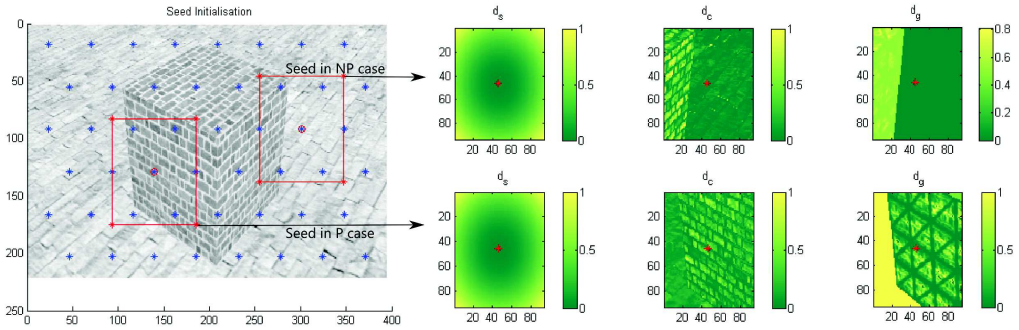

Figure 5: Obtained values for d_{s_0}, d_{c_0} and d_g in two particular cases where the color only criteria can not discriminate pixels. First row: the seed lies on a surface with an unknown geometric distance. Second row: the seed belongs to a surface knowing its orientation, i.e. planar patch, and it aggregates pixels that lies on a surface with the same normal orientation.

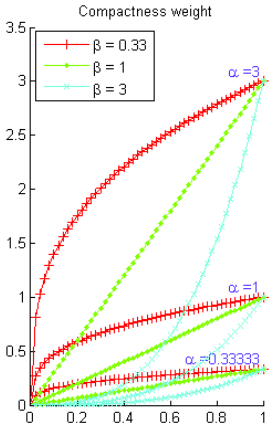

Figure 6: Influence of α and β parameters on the d_{s_0} term related to the compactness.

ness and the two other terms, while β gives a relative importance to the neighbourhood of a given seed.

4 EXPERIMENTATION

For the experiments, the seed initialisation is made on an octagonal structure instead of a regular grid because this shape minimizes the distance between seeds in a neighbourhood.

Preliminary results on a synthetic data with controlled lighting and shape are presented 7. We quantify the quality of the results with two commonly used measures: the boundary recall and the under-segmentation error. The boundary recall measures how well the boundary of the over-segmentation matched with the ground-truth boundary. The under-segmentation error measures how well the set of superpixels described the ground-truth segments.

.

We have remarked that our approach provides compact and geometric coherent superpixels. For a low number of superpixels, when the input parameter K is set to 50 and 100 superpixels, SP_{geom} performs with a higher recall and a lower undersegmentation error than the k-means superpixels approach SP_{5D} tested. Thanks to the geometry information, our method can obtain promising segmentation results.

5 CONCLUSION

In this paper, we have presented a new approach to generate superpixels on calibrated multi-view images by introducing a geometric term in the distance involved in the energy minimization step. This geometric information is a combination of the normal map and the similarity map. Our approach enables to obtain geometric coherent superpixels, i.e. the edges of the superpixels are coherent with the edges of planar patches. The quantitative tests show that the proposed method obtains a better recall and under-segmentation error compared to the k-means approach.

In perspective, we have to generalize this work to real images with meshes that do not respect the edges of the planar surfaces. In order to take into account this drawback, our next algorithm will include a cutting process of the triangles that compose the mesh.

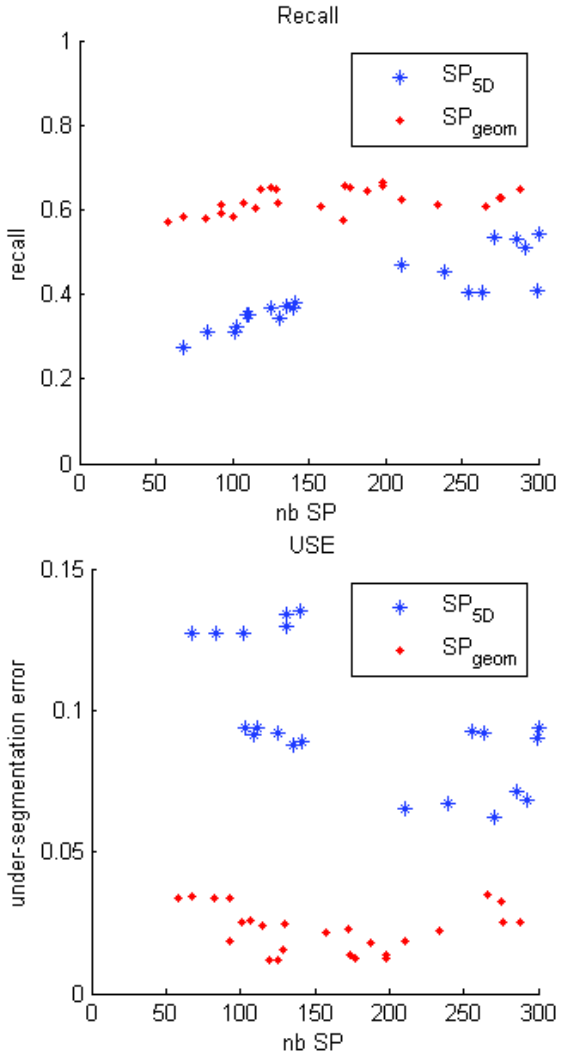

Figure 7: Boundary recall and undersegmenation error for SP_5D based on SLIC and our proposed SP_geom.

REFERENCES

Achanta, R., Shaji, A., Smith, K., Lucchi, A., Fua, P., and Susstrunk, S. (2012). Slic superpixels compared to state-of-the-art superpixel methods.

Arbelaez, P., Maire, M., Fowlkes, C., and Malik, J. (2009). From contours to regions: An empirical evaluation. In *IEEE Computer Vision and Pattern Recognition (CVPR)*.

Author (2015). title.

Bartoli, A. (2007). A random sampling strategy for piecewise planar scene segmentation. In *Computer Vision and Image Understanding (CVIU)*.

Comaniciu, D. and Meer, P. (2002). Mean shift: a robust approach toward feature space analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 24(5).

Felzenszwalb, P. and Huttenlocher, D. (2004). Efficient graph-based image segmentation. In *International Journal of Computer Vision (IJCV)*.

Gallup, D., Frahm, J.-M., and Pollefeys, M. (2010). Piecewise planar and non-planar stereo for urban scene reconstruction. In *IEEE Computer Vision and Pattern Recognition (CVPR)*.

Gould, S., Fulton, R., and Koller, D. (2009). Decomposing a scene into geometric and semantically consistent regions. In *IEEE International Conference on Computer Vision (ICCV)*.

Hartley, R. I. and Zisserman, A. (2004). *Multiple View Geometry in Computer Vision*. Cambridge University Press, ISBN: 0521540518.

Hoiem, D., Efros, A., and Herbert, M. (2005). Geometric context from a single image. In *IEEE International Conference on Computer Vision (ICCV)*.

Levinshtein, A., Stere, A., Kutulakos, K., Fleet, D., Dickinson, S., and Siddiqi, K. (2009). Turbopixels: Fast superpixels using geometric flows. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 31(12):2290–2297.

Mičušík, B. and Košecká, J. (2010). Multi-view superpixel stereo in urban environments. In *International Journal of Computer Vision (IJCV)*.

Moore, A., Prince, S., Warrell, J., Mohammed, U., and Jones, G. (2008). Superpixel lattices. In *IEEE Computer Vision and Pattern Recognition (CVPR)*.

Mori, G. (2005). Guiding model search using segmentation. In *IEEE International Conference on Computer Vision (ICCV)*, volume 2, pages 1417–1423.

Saxena, A., Sun, M., and Ng, A. (2008). Make3d: Depth perception from a single still image. In *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*.

Schick, A., Fischer, M., and Stiefelhagen, R. (2012). Measuring and evaluating the compactness of superpixels. In *ICPR'12*.

Wang, S., Lu, H., Yang, F., and Yang, M. (2011). Superpixel tracking. In *IEEE International Conference on Computer Vision (ICCV)*, pages 1323–1330.

Wang, Z., Bovik, A., Sheikh, H., and Simoncelli, E. (2004). Image quality assessment: From error visibility to structural similarity. In *IEEE TRANSACTIONS ON IMAGE PROCESSING*.

Wu, C. (2011). Visualsfm: A visual structure from motion system.

Z. Wang, Z. and Bovik, A. (2002). A universal image quality index. In *IEEE Signal Processing Letters*.