Time of Flight Image Segmentation through Co-Regularized Spectral Clustering

LUCIANO LORENTI AND JAVIER GIACOMANTONE

Instituto de Investigación en Informática (III-LIDI),
Facultad de Informática - Universidad Nacional de La Plata - Argentina.
La Plata, Buenos Aires, Argentina.
{llorenti, jog, degiusti}@lidi.info.unlp.edu.ar

Abstract. Time of Flight (TOF) cameras generate two simultaneous images, one for intensity and one for range. This allows tackling segmentation problems where the information pertaining to intensity or range alone is not enough to extract objects of interest from a 3D scene. In this paper, we present a spectral segmentation method that combines information from both images. By modifying the affinity matrix of each of the images based on the other, the segmentation of objects in the scene is improved. The proposed method exploits two mechanisms, one for reducing the computational demand when calculating the eigenvectors for each matrix, and another for improving segmentation performance. The experimental results obtained with two sets of real images are presented and used to assess the proposed method.

Keywords: Segmentation, Range Images, Time of Flight Cameras, Spectral Clustering

1. Introduction

Segmentation is generally the first stage in an image analysis system, and it is one of its most critical tasks because it affects subsequent stages [5][21]. Computer vision algorithms, in particular segmentation ones, that have been successfully used in industrial environments, with controlled colors and lighting, do not obtain similar results in different contexts. An alternative to tackle problems in which outline conditions are not suitable for adequate segmentation would be to add depth information, i.e., the distance at which the objects forming the scene are located in relation to the capture device [13][12]. In this context, image segmentation consists in using algorithms that use both sources of information rather than just intensity levels [3][9]. With this perspective, the segmentation problem can be presented as the search for effective shapes that will allow correctly partitioning a set of samples with intensity and distance information. In particular, for this work we used a TOF camera, the MESA SR 4000 [6], that allows obtaining range and intensity images simultaneously. The SR 4000 is an active camera that

uses its own lighting source through a matrix of infrared, amplitude-modulated light emitting diodes. Camera sensors detect the light that is reflected on the illuminated objects and the camera generates two images. The intensity image is proportional to the amplitude of the reflected wave, and the range or distance image is generated from the phase difference between the emitted wave and the wave reflected on each element in the image [2]. The main advantages versus other 3D measurement techniques is the possibility of obtaining images at speeds that are compatible with real-time applications, as well as the possibility of obtaining clouds of 3D points from a single point of view [10][7]. Due to the computational complexity required for spectral clustering algorithms, methods facilitating the calculation of the eigenvectors of the affinity matrix have been proposed recently [11][1][8]. The clustering techniques that have been used allow improving, from multiple data representations, the clustering process [15][14][4]. The method proposed exploits two mechanisms, one for reducing the computational demand when calculating the eigenvectors for each matrix, and another for improving segmentation performance. The improvement in the computational demand is achieved by approximating the eigenvectors of the affinity matrix obtained from each of the images. Segmentation is improved, with respect to the use of a single image, by means of an iteration mechanism that allows obtaining the optimal eigenvector space to carry out the segmentation. The method proposed is assessed by means of two datasets of real images, one obtained with a TOF camera, and another one provided by the Laboratory of Multimedia Technology and Telecommunications of the University of Padua [20].
This article is organized as follows: in Section 2, a review of the main concepts used for the method proposed is presented. The method itself is described in Section 3. In Section 4, experimental results are detailed. Finally, the conclusions are discussed in Section 5.

2. Spectral clustering

Given a set of patterns $X = \{x_1, x_2, ...x_n\} \in \mathbb{R}^m$, and a similarity function $d : \mathbb{R}^m \times \mathbb{R}^m \to \mathbb{R}$, an affinity matrix W can be built so that $W(i,j) = d(x_i, x_j)$. Spectral clustering algorithms obtain a representation of the data in a lower dimension space by solving the following optimization problem:

$$\begin{aligned} \max_{U \in \mathbb{R}^{n \times k}} \quad & Tr\left(U^T L U\right) \\ \text{s.t.} \quad & U^T U = I \end{aligned} \tag{1}$$

where $L = D^{-\frac{1}{2}} W D^{-\frac{1}{2}}$ is the Laplacian matrix of W in accordance to [17] and D is a diagonal matrix with the sum of the rows in W on its main diagonal. After U is obtained, its rows are considered as the new coordinates

for the patterns. With this new representation, traditional clustering algorithms are easier to apply [19].
Image segmentation spectral methods are based on the eigenvectors and eigenvalues of a N x N matrix derived from the affinities among pixels. It should be noted that one of the main limitations of this type of algorithms is the amount of memory required due to the fact that the dimensions of W exhibit quadratic growth with respect to the number of elements in the image. A possible approach to deal with this problem is using a sparse matrix that codes the local information from each pixel. In this representation, each element is connected only to some of its closest neighbors, and all other connections are assumed to be zero [19] [16]. Alternatively, the affinities of a small set of pixels can be calculated and the remaining affinities, approximated [11] [1].

2.1 Approximate Calculation of Eigenvectors

One of the initial proposals to define spectral clustering relates the weights matrix W to a graph incidence matrix and to the clustering problem as a graph partitioning problem [19]. From this perspective, each of the X_i patterns is considered as the vertex of a weighted, non-directed graph $G = (V, E)$ and the element $W(i, j)$ is the weight of the edge that connect vertex i with vertex j. Be $G = (V, E)$ the similarity graph derived from a set of patterns $X = \{x_1, x_2, ..., x_n\}$, $A \subset V$ a subset of sampled vertexes and $B = V - A$, the remaining, non-sampled vertexes. G_A is the graph that results from connecting the vertexes of A to each other, and G_B is the graph that results from connecting the vertexes of A to the vertexes of B. Be W_A the adjacency matrix of G_A and L_A the Laplacian matrix of G_A. Be W_B and L_B the corresponding matrixes of G_B. We can then formulate the adjacency matrix of G, which we will call W, and the Laplacian matrix of G, which we will call L, as follows:

$$W = \begin{bmatrix} W_A & W_B \\ W_B^T & W_C \end{bmatrix} \qquad L = \begin{bmatrix} L_A & L_B \\ L_B^T & L_C \end{bmatrix}$$

Considering the diagonalization of $A = U \Lambda U^T$, using Nystrom's extension [11]: $\overline{U} = \begin{bmatrix} U \\ B^T U \Lambda^{-1} \end{bmatrix}$ as approximate eigenvectors of W, an approximation of W, called $\hat{W}$, can be obtained by calculating only A and B:

$$\hat{W} = \bar{U} \Lambda \bar{U}^T = \begin{bmatrix} A & B \\ B^T & B^T A^{-1} B \end{bmatrix}$$

To obtain the eigenvectors for $\hat{L} = \hat{D}^{\frac{1}{2}}\hat{W}\hat{D}^{\frac{1}{2}}$, i.e. the approximate Laplacian matrix generated from $\hat{W}$ the proposed technique can be used, which only requires calculating $\hat{L_A}$ and $\hat{L_B}$:

$$\hat{L_{Aij}} = \frac{W_{A_{ij}}}{\sqrt{\hat{d}_i \hat{d}_j}} \quad \hat{L_{Bij}} = \frac{W_{B_{ij}}}{\sqrt{\hat{d}_i \hat{d}_{j+|A|}}} \tag{2}$$

where $\hat{d} = \hat{W}\mathbf{1} = \begin{bmatrix} a_r + b_r \\ b_c + B^T A^{-1} b_r \end{bmatrix}$ and a_r represent the sum of the rows in A, b_c represents the sum of the columns in B y b_r represents the sum of the rows in B. If $\hat{L_A}$ is defined positive, the approximate orthogonal eigenvectors can be found in a single step. Be $S = \hat{L_A} + \hat{L_A}^{-\frac{1}{2}} \hat{L_B} \hat{L_B}^T \hat{L_A}^{-\frac{1}{2}}$ and its diagonalization $S = U_S \Lambda_S U_S^T$, Fowkles et al [11] showed that if a matrix V if defined as:

$$V = \begin{bmatrix} \hat{L_A} \\ \hat{L_B}^T \end{bmatrix} \hat{L_A}^{-\frac{1}{2}} U_S \Lambda_S^{-\frac{1}{2}} \tag{3}$$

$\hat{LN}$ is diagonalized by V and by Λ_S and $V^T V = I$

2.2 Co-regularization

When the dataset has more than one representation, these representations are referred to as "views". In the context of spectral clustering, co-regularization techniques help support similarities in the examples in the new representation generated from the eigenvectors of each of the views. Let $X^{(v)} = \{x_1^{(v)}, x_2^{(v)}, ..., x_n^{(v)}\}$ be the examples for view V y $L^{(v)}$ the Laplacian matrix created from X for view V. We define $U^{(v)}$ as the matrix formed by the first k eigenvectors corresponding to matrix $L^{(v)}$ accordance with (1).

In [14] a criterion to measure the dissimilarity degree between two representations was proposed:

$$D(U^{(v)}, U^{(w)}) = \left|\left| \frac{K_{U^{(v)}}}{||K_{U^{(v)}}||_F} - \frac{K_{U^{(w)}}}{||K_{U^{(w)}}||_F} \right|\right|_F^2$$

Where $K_{U^{(v)}}$ is the similarity matrix generated from the patterns in the new representation $U^{(v)}$ and $||\cdot||_F$ is Frobenius norm. If the inner product between the vectors is used as a similarity measure, we obtain

$K_{U^{(v)}} = U^{(v)}U^{(v)^T}$. Ignoring additive and escalation constants, the previous equation can be formulated as follows:

$$D(U^{(v)}, U^{(w)}) = -Tr\left(U^{(v)}U^{(v)^T}U^{(w)}U^{(w)^T}\right)$$

The goal is minimizing the level of disagreement among the representations obtained from each view. Therefore, the following optimization problem is obtained that combines the individual objectives of spectral clustering and the objective that determines the disagreement among representations:

$$\begin{aligned} \max_{\substack{U^{(v)} \in R^{n\times k} \\ U^{(w)} \in R^{n\times k}}} \quad & Tr\left(U^{(v)^T}L^{(v)}U^{(v)}\right) + Tr\left(U^{(w)^T}L^{(w)}U^{(w)}\right) + \\ & \lambda Tr\left(U^{(v)}U^{(v)^T}U^{(w)}U^{(w)^T}\right) \\ \text{s.t.} \quad & U^{(v)^T}U^{(v)} = I \\ & U^{(w)^T}U^{(w)} = I \end{aligned} \tag{4}$$

The parameter λ balances the objective of spectral clustering and the disagreement among representations. The joint optimization problem can be solved by using alternating maximization with respect $U^{(v)}$ and $U^{(w)}$. For a given $U^{(w)}$ the following optimization problem is obtained for $U^{(v)}$:

$$\begin{aligned} \max_{U^{(v)}\in R^{n\times k}} \quad & Tr\left(U^{(v)^T}\left(L^{(v)} + \lambda U^{(w)}U^{(w)^T}\right)U^{(v)}\right) \\ \text{s.t.} \quad & U^{(w)^T}U^{(w)} = I \end{aligned} \tag{5}$$

Which results in a traditional clustering algorithm with the modified Laplacian matrix $L^{(v)} + \lambda U^{(w)}U^{(w)^T}$

3. Method Proposed

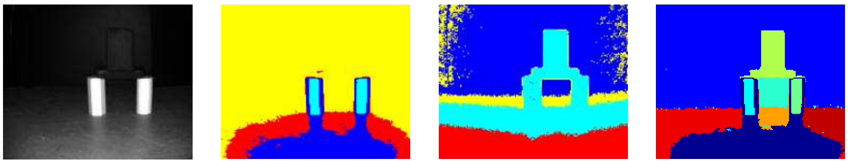
Let I be an amplitude image and be R a dimension range image $n\times m$, both for the same scene.

1. From I and R, the approximate Laplacian matrixes $\hat{L_I}$ and $\hat{L_R}$ are obtained, as described in (2).
2. Let $\hat{V_I}$ be the approximate eigenvectors for $\hat{L_I}$ calculated in accordance with (3).
3. $\hat{V_R}$, the eigenvectors of the modified Laplacian matrix $\hat{L_R} + \lambda\hat{V_I}\hat{V_I}^T$ (5) are obtained using method (2).

4. V_I, the eigenvectors of the modified Laplacian matrix $\hat{L}_I + \lambda \hat{V}_R \hat{V}_R^T$ (5) are obtained using method (2).
5. $V = [V_I \; V_R]$
6. A clustering algorithm is applied to V.
7. The criterion proposed in [8] is used to assess segmentation algorithm performance. If performance improves, then go to 3; if not, end.

4. Experimental Results

The performance of the segmentation algorithm proposed was assessed over 50 captured images using the MESA SwissRanger SR4000 time-of-flight camera [6] and the entire dataset provided by the Laboratory of Multimedia Technology and Telecommunications of the University of Padua [20].



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

Figure 1. Segmentation applied to an image captured with the SwissRanger SR4000 time-of-flight camera. (a) Amplitude image of a real scene. (b) [11] method applied to the amplitude image. H=0.24 (c) [11] method applied to the range image. H=0.18 (d) Proposed method $\lambda = 3$. H=0.29.

The MESA SwissRanger SR4000 time-of-flight camera provides two images: an amplitude image and a range image, both with 144 x 176 pixels. The dataset [20] contains images captured with a time-of-flight camera and a traditional RGB camera. The performance of the segmentation algorithm was assessed using the criterion proposed in [8] and [18], which we called H.
The similarity function used in every case takes into account the spatial distribution of pixels in the image and the difference between their values:

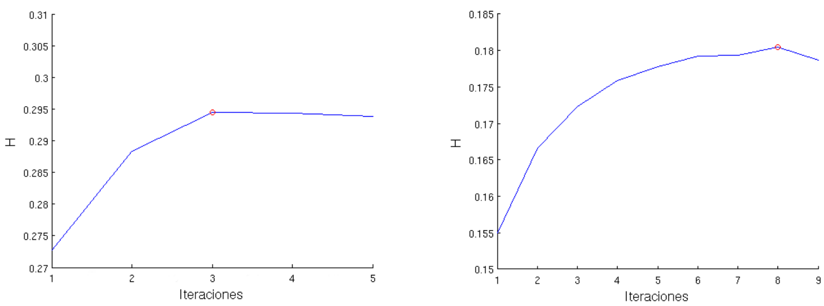
$$W(i,j) = e^{-\frac{\|X_{(i)} - X_{(j)}\|_1^2}{sX}} * e^{-\frac{\|F_{(i)} - F_{(j)}\|_1^2}{sY}}$$

where $X(i)$ is the spatial location of pixel i, $F(i)$ is the value of the i-th pixel in the image, $sX = E(\|X_{(i)} - X_{(j)}\|_1^2) + \frac{3}{4}\sigma(\|X_{(i)} - X_{(j)}\|_1^2)$ of the pixels in set A and $sY = E(\|F_{(i)} - F_{(j)}\|_1^2)$ of the pixels in set A. Figure 1 shows the experimental results of the method proposed applied to an image obtained with the MESA SwissRanger SR4000 time-of-flight camera. The amplitude image of the captured scene 1(a) presents 3 objects over a black background, all at the same distance. One of the objects has an intensity level that is similar to that of the background, which makes segmentation difficult. Since all objects at the front of the image are at the same distance, their range values will be similar. Figures 1(b) and 1(c) show the result of applying method [11] to the amplitude image and the range image, respectively. Figure 1(d) shows the result of applying the proposed method at the optimal operation point. The method correctly combines the information from both noisy images to segment the objects found in the scene. Figure 3(a) shows the performance assessed for each iteration of the algorithm. Figure 2 shows the result of applying the proposed algorithm to a scene in the dataset provided by the University of Padua. Figure 2(a) shows the amplitude image for the scene. Figures 2(b) and 2(c) show the result of applying algorithm [11] to the amplitude and range images. Separately, both images do not provide the necessary information to extract all objects in the scene.

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

Figure 2. Segmentation of an image of the dataset provided by the University of Padua. (a) Amplitude image of a real scene (b) H=0.11 (c) H=0.13 (d) Proposed method using $\lambda = 3$. H=0.18.

The proposed method, through co-regularization, is successful in extracting the useful information from both images, maximizing segmentation performance, as Figure 2(d) shows. The performance assessed in each iteration of the algorithm is shown in Figure 3(b).



(a) (b)

Figure 3. Performance in relation to the number of iterations

5. Conclusions

In this article, we presented a clustering method for segmenting images captured by TOF cameras. The preliminary results obtained both on intensity and range images are satisfactory. The algorithm correctly combines the information provided by both images, even in the presence of noise, by using co-regularization techniques. The performance obtained was better when using semi-supervised learning instead of using concatenated characteristics in all tested cases.

In a future stage, we plan adding color information to the segmentation algorithm. Also, we will consider the convenience of using an alternative disparity measure.

Bibliography

1. Nicholas Arcolano. *Approximation of positive semidefinite matrices using the Nyström method*. PhD thesis, Harvard University, 2011.
2. N. Blanc, T. Oggier, G. Gruener, J. Weingarten, A Codourey, and P. Seitz. Miniaturized smart cameras for 3d-imaging in real-time [mobile robot applications]. In *Sensors, 2004. Proceedings of IEEE*, pages 471-474 vol.1, Oct 2004.
3. Amit Bleiweiss and Michael Werman. Fusing time-of-flight depth and color for real-time segmentation and tracking. In *Dynamic 3D Imaging*, pages 58-69. Springer, 2009.
4. Hanneke Borgdorff, Evgeni Tsivtsivadze, Rita Verhelst, Massimo Marzorati, Suzanne Jurriaans, Gilles F Ndayisaba, Frank H Schuren, and Janneke HHM van de Wijgert. Lactobacillus-dominated cervicovaginal microbiota associated with reduced hiv/sti prevalence and genital hiv viral load in african women. Nature Publishing Group, 2014.
5. John Canny. A computational approach to edge detection. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, PAMI-8(6):679-698, Nov 1986.
6. Miguel Cazorla, Diego Viejo, and Cristina Pomares. Study of the sr 4000 camera. In *XI Workshop de Agentes FÃ□sicos*, 2004.
7. F. Chiabrando, D. Piatti, and F. Rinaudo. R-4000 tof camera: Further experimental tests and first applications to metric surveys. In *Proceedings of V Symposium on Remote Sensing and Spatial Information Sciences*, pages 149-154, 2010.
8. C. Dal Mutto, P. Zanuttigh, and G.M. Cortelazzo. Fusion of geometry and color information for scene segmentation. *Selected Topics in Signal Processing, IEEE Journal of*, 6(5):505-521, Sept 2012.
9. G. Danciu, M. Ivanovici, and V. Buzuloiu. Improved contours for tof cameras based on vicinity logic operations. In *Optimization of Electrical and Electronic Equipment (OPTIM), 2010 12th International Conference on*, pages 989-992, May 2010.
10. Adrian A Dorrington, Cameron DB Kelly, Shane H McClure, Andrew D Payne, and Michael J Cree. Advantages of 3d time-of-flight range imaging cameras in machine vision applications. In *Proceedings of 16th Electronics New Zealand Conference*, pages 95-99, 2009.

11. Charless Fowlkes, Serge Belongie, Fan Chung, and Jitendra Malik. Spectral grouping using the nyström method. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(2):214-225, Feburary 2004.
12. Guruprasad M Hegde and Cang Ye. A recursive planar feature extraction method for 3d range data segmentation. In *Systems, Man, and Cybernetics (SMC), 2011 IEEE International Conference on*, pages 3119-3124. IEEE, 2011.
13. Dirk Holz, Stefan Holzer, Radu Bogdan Rusu, and Sven Behnke. Real-time plane segmentation using rgb-d cameras. In *RoboCup 2011: Robot Soccer World Cup XV*, pages 306-317. Springer, 2012.
14. Abhishek Kumar, Piyush Rai, and Hal Daume. Co-regularized multi-view spectral clustering. In J. Shawe-Taylor, R.S. Zemel, P.L. Bartlett, F. Pereira, and K.Q. Weinberger, editors, *Advances in Neural Information Processing Systems 24*, pages 1413-1421. Curran Associates, Inc., 2011.
15. Abhishek Kumar, Avishek Saha, and Hal Daume. Co-regularization based semi-supervised domain adaptation. In *Advances in Neural Information Processing Systems*, pages 478-486, 2010.
16. Luciano Lorenti and Javier Giacomantone. Segmentación espectral de imágenes utilizando cámaras de tiempo de vuelo. In *XVIII Congreso Argentino de Ciencias de la Computación*, 2013.
17. Andrew Y. Ng, Michael I. Jordan, and Yair Weiss. On spectral clustering: Analysis and an algorithm. In *ADVANCES IN NEURAL INFORMATION PROCESSING SYSTEMS*, pages 849-856. MIT Press, 2001.
18. C. Rosenberger and K. Chehdi. Genetic fusion: application to multi-components image segmentation. In *Acoustics, Speech, and Signal Processing, 2000. ICASSP '00. Proceedings. 2000 IEEE International Conference on*, volume 6, pages 2223-2226 vol.4, 2000.
19. J. Shi and J. Malik. Normalized cuts and image segmentation. In *Computer Vision and Pattern Recognition, 1997. Proceedings, 1997 IEEE Computer Society Conference on*, pages 731-737, Jun 1997.
20. Multimedia Technology and University of Padova Telecommunications Laboratory. Joint color and depth segmentation datasets, July 2014.
21. R. Wu, Z. y Leahy. An optimal graph theoretic approach to data clustering: theory and its application to image segmentation. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 15(11):1101-1113, Nov 1993.

Weinberger, editors, *Advances in Neural Information Processing Systems 24*, pages 1413-1421. Curran Associates, Inc., 2011.
15. Abhishek Kumar, Avishek Saha, and Hal Daume. Co-regularization based semi-supervised domain adaptation. In *Advances in Neural Information Processing Systems*, pages 478-486, 2010.
16. Luciano Lorenti and Javier Giacomantone. Segmentación espectral de imágenes utilizando cámaras de tiempo de vuelo. In *XVIII Congreso Argentino de Ciencias de la Computación*, 2013.
17. Andrew Y. Ng, Michael I. Jordan, and Yair Weiss. On spectral clustering: Analysis and an algorithm. In *ADVANCES IN NEURAL INFORMATION PROCESSING SYSTEMS*, pages 849-856. MIT Press, 2001.
18. C. Rosenberger and K. Chehdi. Genetic fusion: application to multi-components image segmentation. In *Acoustics, Speech, and Signal Processing, 2000. ICASSP '00. Proceedings. 2000 IEEE International Conference on*, volume 6, pages 2223-2226 vol.4, 2000.
19. J. Shi and J. Malik. Normalized cuts and image segmentation. In *Computer Vision and Pattern Recognition, 1997. Proceedings, 1997 IEEE Computer Society Conference on*, pages 731-737, Jun 1997.
20. Multimedia Technology and University of Padova Telecommunications Laboratory. Joint color and depth segmentation datasets, July 2014.
21. R. Wu, Z. y Leahy. An optimal graph theoretic approach to data clustering: theory and its application to image segmentation. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 15(11):1101-1113, Nov 1993.