MULTIPLE VIEW REGION MATCHING AS A LAGRANGIAN OPTIMIZATION PROBLEM

Felipe Calderero, Ferran Marqués

Department of Signal Theory and Communications
Technical University of Catalonia (UPC)
Barcelona, Spain

Antonio Ortega

Department of Electrical Engineering
University of Southern California
Los Angeles, California, USA

ABSTRACT

A method to establish correspondences between regions belonging to independent segmentations of multiple views of a scene is presented. The trade-off between color similarity and projective similarity of the matching regions is formulated in terms of a constrained optimization, analogous to a Rate-Distortion budget-constrained allocation problem, and solved using Lagrangian optimization techniques.

Index Terms— Multiple view, region matching, Lagrangian optimization, image segmentation

1. INTRODUCTION

The idea behind this work is the possibility to exploit the information in multiple views of a 3D scene to perform a joint segmentation among them. As images belong to the same 3D scene, the resulting partitions will take into account spatial and depth information, outperforming an independent 2D segmentation of each view in terms of similitude with a direct segmentation of the 3D space.

In general, image or scene segmentation is an ill-posed problem. Obviously, the best segmentation fitting the data is the own data, that is, every pixel an isolated region. Hence, a regularization term is needed. A typical approach is to perform a hierarchical segmentation; that is, a segmentation through various steps relying on different criteria of increasing complexity, using features such as color, contour complexity, or texture [1].

In a multiple view scenario, the regularization can be handled not only from features on the image but also with the spatial or depth information carried by all the views. Generally speaking, the key point of this approach is to regularize the segmentation in one of the views using the set of remaining views. This way, the cost of merging a region into an image is related with the cost of merging the correspondent region into the 3D scene under a spatial criterion.

Nevertheless, before taking advantage of complex 3D information within the segmentation process, the correspondences among regions in different views must be established. A region matching is performed; that is, each region on a view is associated, where possible, to its equivalent regions in the other views.

Similarly, region matching techniques have been used in motion estimation and motion-based segmentation [2], finding correspondences between regions in consecutive frames. They have proved to be also useful in stereoscopic vision, although they have been usually applied not for segmentation purposes, but for dense disparity map estimation of pairs of views [3].

This work has been partly supported by the EU project NoE SIMILAR FP6-507609 and by the grant TEC2004-01914 of the Spanish Government.

In general, due to occlusions or to the finite dimensions of images, a region appearing in a view may not have its correspondent one either in some of the views or none of them. In addition, depending on the segmentation process, a region could partially match another region or have multiple correspondences.

The work presented in this paper tackles the initial step on the multiple view region matching problem. Following the previous idea of a segmentation based on different criteria of increasing complexity, we propose a method to establish region correspondences in different views, starting from an initial fine partition for each view, obtained independently under some criteria (for instance, color). Since the proposed method is the initial 3D step in the multiple view segmentation process, the criterion to be used is still simple (combining color similarity and spatial coherence). Moreover, the result should not be a complete matching of all regions in the different partitions but a partial matching of the most reliable regions, which should allow the robust estimation of more complex criteria for the subsequent matching and merging steps.

The method is based on a constrained optimization technique, inspired on the Rate-Distortion (R-D) Optimization Theory. The idea is to reduce the set of possible matching regions taking into account color similarity constrained to spatial coherence among them, and take advantage of the Lagrangian optimization techniques to solve the problem. At this initial stage the method is partition dependent, i.e. matching results depend on the selected partition to match and, thus, the established correspondences may not be symmetric with respect to the other possible partition selections.

In the next section, the multiple view region matching problem is formally stated. Similarity measures in both color and projective spaces are defined in Section 3. In Section 4 the multiple view region matching problem is formulated as a constrained optimization problem and solved using Lagrangian optimization techniques. The method is applied to a set of multiple views in Section 5. Finally, conclusions are presented in Section 6.

2. MULTIVIEW REGION MATCHING PROBLEM

Formally, our region matching problem is stated as follow:

Problem Definition. *Let us assume that we have N views of the same scene, $V^1 ... V^N$, with partitions, $\Pi^1 ... \Pi^N$. These partitions are obtained independently for each view. Without loss of generality, let us select the partition of the first view, Π^1. Our goal is to determine, for each region of Π^1, $\rho_i^1 \in \Pi^1$, which regions in the remaining partitions better match ρ_i^1, taking into account that a matching may not exist for some views.*

To allow unmatched regions is equivalent to adding an *empty*

region to the set of regions in each partition. For simplicity, let us define the extended set of partitions as:

$$\Gamma^k : \{\emptyset, \rho_i^k \in \Pi^k\} \tag{1}$$

Thus, the correspondences of each region can be compacted into a matching array:

$$\rho_i^1 \in \Pi^1 \Rightarrow \underline{\gamma}_i = [\gamma_i^2 \dots \gamma_i^N], \quad \gamma_i^k \in \Gamma^k \tag{2}$$

Note that this formulation performs a *many-to-one* mapping into a given Γ^k (neither injective nor surjective). The possibility of multiple correspondences (*many-to-many*), that is not addressed in this paper, could be interesting in the case of having oversegmented views.

In addition, it cannot be assured that the identified matchings will be symmetric, i.e $\rho_i^1 \Rightarrow \rho_j^k$ does not imply that $\rho_j^k \Rightarrow \rho_i^1$. This can be simply addressed by performing the matching when selecting each partition and then removing those matchings that are no symmetric.

3. SIMILARITY CRITERIA

3.1. Color Similarity

A first matching criterion is color similarity between regions in different views. Let us assume that the color distribution of a region can be modeled as a three-dimensional (RGB) spatially independent normal distribution, $N(\mu_\rho, \sigma_\rho)$, abbreviated as N_ρ, where $\mu_\rho, \sigma_\rho \in \mathbb{R}^3$ are the mean and variance of the color components of pixels belonging to the region. Hence, the similarity measure between region colors can be estimated as a distance between statistical distributions using, for instance, the *Kullback-Leibler* (KL) divergence [4]:

$$d_{KL}(p,q) = \int_{-\infty}^{+\infty} q(x) \log \frac{q(x)}{p(x)} dx \tag{3}$$

We choose a symmetric version of the KL distance, known as *Jeffreys* (JF) distance [5] as color similarity measure:

$$d_{JF}(p,q) = \frac{1}{2}\left[d_{KL}(p,q) + d_{KL}(q,p)\right] \tag{4}$$

Thus, the color distance between two regions is estimated as the JF distance of their color distributions, assuming they are Gaussian. The JF distance can be generalized to determine the color distance of N different regions as follows:

$$d_{color}(\rho_1, \cdots, \rho_N) = \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} d_{JF}(N_{\rho_i}, N_{\rho_j}) \tag{5}$$

The JF distance is chosen for its simplicity for the Gaussian case, and because it allows a correct comparison even when the color distributions of two regions only match partially (for instance, when a region is embedded in the other) [6].

3.2. Projective Similarity

The projective similarity between two regions will be based on the *epipolar distance* between their centroids. Formally, the epipolar distance from a point in a view, x^1, with respect to a point in another view, x^2, is defined as the Euclidean distance from the epipolar line generated by x^1 into the second view, $l^2_{x^1}$, to the point x^2:

$$d_{Epi}(x^1, x^2) = d_{Euclidean}(l^2_{x^1}, x^2) \tag{6}$$

To obtain a symmetric distance function, as before, we can combine distances computed in both directions. The symmetric measure we will use is known as *symmetric epipolar* (SE) distance [7]:

$$d_{SE}(x^1, x^2) = \sqrt{d^2_{Epi}(x^1, x^2) + d^2_{Epi}(x^2, x^1))} \tag{7}$$

Its generalization for N views can be written as:

$$d_{SE}(x^1, \dots, x^N) = \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} d_{SE}(x^i, x^j) \tag{8}$$

Thus, the projective similarity between a set of regions in N different views, $\rho_i^1 \dots \rho_j^n \dots \rho_k^N$, is defined as the SE distance of the set of their centroids, $c_i^1 \dots c_j^n \dots c_k^N$:

$$d_{proj}(\rho_i^1 \dots \rho_j^n \dots \rho_k^N) = d_{SE}(c_i^1 \dots c_j^n \dots c_k^N) \tag{9}$$

In this case, projective similarity is understood as measuring how accurately the region centroids can represent projections of the same 3D point. This measure is simple and fast to compute, unlike other proposed distances where inverse matrix computations or homography estimation are required [8].

4. MULTIVIEW REGION MATCHING ALGORITHM

4.1. Rate-Distortion Analogy

Once similarity criteria have been defined, we have to determine the set of regions that provides the best compromise between both cost functions. We tackle this problem using a constrained optimization framework. Intuitively, we could think of minimizing both: color and projective distance. Nevertheless, minimizing the projective distance may not lead to the best solution, since the matched region centroids may not backproject to the same 3D point, although we expect them to be relatively close.

Symmetric epipolar distance penalizes regions far from the epipolar line. Thus, we can determine a maximum allowed value for the SE distance and to define a reasonable projective area into the other views where a particular region can match its most color similar region.

Taking these last considerations into account, the problem stated in Section 2, i.e., finding the correspondences for each region of a selected partition in the other partitions, can be formulated as an optimization problem, analogous to the budget-constrained allocation problem in video coding [9].

Let us define the total color distance and the total projective distance of a partition matching, respectively, as the sum of the distance of each region with respect to its array of matching regions in each view, $\underline{\gamma}_i$, as defined in Section 3:

$$D_{color} = \sum_{i=1}^{\text{Card}(\Pi^1)} d_{color}(\rho_i^1, \underline{\gamma}_i) \tag{10}$$

$$R_{proj} = \sum_{i=1}^{\text{Card}(\Pi^1)} d_{proj}(\rho_i^1, \underline{\gamma}_i) \tag{11}$$

where $\text{Card}(\Pi^1)$ is the cardinality (number of regions) of Π^1.

Following the analogy with R-D formulation for video coding, D_{color} can be understood as the total *distortion* to minimize, and R_{proj} as the total *rate* available. Hence, the problem can be written:

$$\min_{\underline{\gamma}_1 \cdots \underline{\gamma}_N} D_{color} \quad \text{s.t.} \quad R_{proj} \leq R_{proj}^{Budget} \tag{12}$$

Fig. 1: Example of multiple view region matching. The selected region in the first partition (yellow region on left image) is matched to a region into the second and third view partitions. The epipolar lines generated by the region centroids in the other views are plotted (green). The parameter setting for this example is: $\lambda = 0.3$, $\beta = 3$, where λ was determined by the method proposed in Section 4.3.

where R_{proj}^{Budget} is the total projective error accepted in the complete matching process and it is estimated as the sum of the security areas defined for each partition (maximum estimated projective distance at which a region can match).

Thanks to this formulation, we can solve the region matching problem applying the R-D optimization techniques. Given that the trade-off between color and projective distance is identical to the rate-distortion trade-off, i.e. increasing projective distance will always lead to lower or equal color distance, R-D methods can be used to determine operational points on the convex hull of the characteristic. As proved in [10], this problem is equivalent to the Lagrangian optimization problem:

$$\min_{\underline{\gamma}_1 \cdots \underline{\gamma}_N} D_{color} + \lambda \cdot R_{proj}, \quad \lambda \geq 0 \tag{13}$$

for the particular case $R_{proj} \leq R_{proj}^{Budget}$ and its solution is also the optimal solution of (12).

Since the set of correspondences for a partition may be found independently for each region as a consequence of (10) and (11), the sum is obviously minimized by simply minimizing:

$$\sum_{i=1}^{\text{Card}(\Pi^1)} \min_{\underline{\gamma}_i} [d_{color}(\rho_i^1, \underline{\gamma}_i) + \lambda \cdot r_{proj}(\rho_i^1, \underline{\gamma}_i)] \tag{14}$$

Note that this method performs a joint optimization, taking into account the whole set of views, while implicitly handles the possibility that a region may not have a correspondence (i.e., when it matches the empty set) into some of the other views, thanks to considering the extended set of partitions, Γ^k.

To close this formulation, there are two remaining issues. First, the color distance and the projective distance with respect to the empty set need to be defined (Section 4.2). Second, a method to select λ has to be defined (Section 4.3).

4.2. Distance from a Region to the Empty Set

We need to define both, color distance and projective distance between a region and the empty set, $d_{color}(\rho_i, \emptyset)$ and $d_{proj}(\rho_i, \emptyset)$, respectively. In other words, which color distance (*distortion*) is assigned to a region that does not match any regions into another view, and what cost in terms of projective distance (*rate*) this causes.

Our method has to avoid matching a region when its trade-off between color and projective similarity is not acceptable. Only pairs leading to a low projective distance should match. On the contrary, if all possible pairs for a given region lead to a prohibitively large amount of the total projective budget, the region should be assigned to the empty set. Consequently, unmatched regions should be associated to very low color distances (*distortion*) and to very large projective distances (*rate*). Arbitrarily, we can define $d_{color}(\rho_i, \emptyset) = 0$.

A large projective distance for a region will be related with the maximum distance where we expect the region to match. Determining this value exactly will depend on the position of each view's camera and on the specific area and shape of the region. To simplify the problem, we assume that the pixels of a region are uniformly distributed around its centroid (in other words, regions are approximately circular). This way the maximum matching distance for a region can be specified in terms of its mean radium, i.e. the mean distance from the region centroid to its border, $\overline{r}_\rho = \sqrt{A_\rho/\pi}$. The projective distance assigned to an unmatched region should be larger than the expected maximum distance for that region to match, and also proportional to its mean radium:

$$d_{proj}(\rho_i, \emptyset) = \beta \cdot \overline{r}_{\rho_i} \tag{15}$$

This definition takes into account that large regions have a high probability to be matched in other views, and consequently, not matching them has a larger penalty. The parameter β controls the number of regions finding a correspondence or remaining unmatched; equivalently, how expensive an unmatched region is in terms of rate.

4.3. Determining the Value of λ

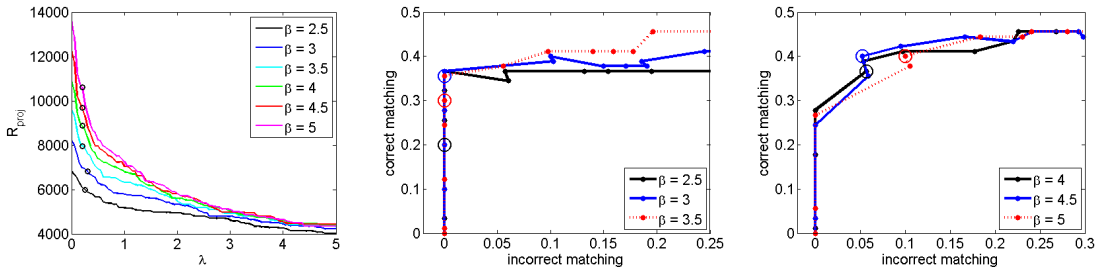
In the video coding framework, solving the R-D problem via Lagrangian optimization is equivalent to finding the value of λ that provides the operational point with minimum total distortion, for a specified total rate; or viceversa, the minimum rate for a given distortion.

On the contrary, in multiple view region matching, we are not interested in specifying a budget in terms of total projective distance. Instead of allocating a finite amount of resources, our goal is to optimally combine two different matching criteria.

Thus, we will redefine the optimization in terms of the probability of incorrect matching (*false alarm*) and probability of correct matching (*hit*). We would like to determine the value of λ that provides the maximum probability of correct matching for a specified probability of incorrect matching; or viceversa, the minimum probability of incorrect matching given a probability of correct matching.

For this purpose, instead of a R-D curve, we will build an *operating characteristic* as a function of λ, identical to the operating curve of a classifier: vary the λ parameter and plot the resulting correct and incorrect matching rates. Note that, in general, this curve will not be either symmetric or concave, since both probabilities will correspond to arbitrary (non Gaussian) multidimensional distributions [11]. However, $\lambda \rightarrow 0$ implies a vanishing probability of false alarm (see Fig.2).

In order to automatically select a λ value, we propose to analyze the set of projective distance (*rate*) curves, i.e. the curves generated by (11) when varying λ. As stated in the Introduction, we want to apply in the first steps of the 3D segmentation a very conservative policy. Thus, we select the inflection point, related to the



(a) Projective distance curves (b) Operating characteristics (c) Operating characteristics

Fig. 2: (a) Projective distance (*rate*) curve for different values of β, as a function λ, when left partition in Fig.1 was selected and matched to the other two partitions. The black circles show the value of λ corresponding to the inflection point, related with the first maximum of the first derivative, of a low-pass version of the projective distance curve. (b)(c) Operating characteristics (computed using the groundtruth) for different values of β, as a function of λ. The circles show which point of the operating curve corresponds to the proposed λ selection in (a).

first maximum of the first derivative, of a low-pass version of the projective distance curve (see Fig.2). This point corresponds approximatively with the end of the fast descending initial part of the projective distance curve, and provides a good estimate of an operational point with null or low incorrect matching rate and the highest correct matching rate (see Fig.2).

5. RESULTS

The results shown in this section are obtained from the set of synthetic views shown in Fig.1, selected for the large amount of ambiguity between regions, specially on the floor. The corresponding partitions for each view (superposed in black in Fig.1) were generated by a color-based region merging procedure [1] until only 100 regions remained. Figure 1 shows an example where the correspondences to the other views for a selected region in the first partition were correctly found.

The projective distance curves, as a function of λ, are plotted in Fig.2 for different values of the projective cost assigned to an unmatched region (varying β, see Section 4.2). Their corresponding operating characteristics, computed using the region matching groundtruth, are also shown. The number of matching errors was reduced computing the matchings when selecting each partition and keeping only those coinciding at least in two partitions.

In Fig.2 the proposed selection method for λ is illustrated. For a low-pass filtered version of the projective distance curve, the value of λ corresponding to the inflection point related to the first maximum of the first derivative is determined. Observing the operational characteristics, it can be seen that this point provides a conservative choice, with a null or low incorrect matching rate while yielding a high correct matching rate. Note that up to a 30% of all possible correct matchings are determined for a null probability of incorrect matching which is a very high value for an initial matching step not using complex criteria (e.g.: shape or neighborhood information).

6. CONCLUSIONS

We have presented a multiple view region matching method based on the similarity between regions in both color and projective space. Although these measures are relatively simple, they correctly characterize the correspondences between regions without any knowledge of the shape of the regions, and therefore, they can be used in an early stage in a hierarchical 3D segmentation.

To approach the stated problem we have taken advantage of a formulation of the problem as a constrained optimization and we have solved it using Lagrangian optimization techniques.

Once specified how expensive is not to match a region, i.e. controlling the number of found correspondences, we propose a way to estimate a conservative value of λ, i.e. providing the highest number of correct matchings for a null or low incorrect matching rate.

Current work aims first at allowing a region in the selected partition matching more than one region in the remaining partitions. Moreover, the addition of contour and neighborhood information is being analyzed to implement the subsequent steps of the hierarchical multiple view segmentation.

7. REFERENCES

[1] P. Salembier and F. Marqués, "Region-based representations of image and video: Segmentation tools for multimedia services," *IEEE Trans. Circuits and Systems for Video Technology*, vol. 9, no. 8, pp. 1147–1167, 1999.

[2] S.S. Beauchemin and J.L. Barron, "The computation of optical flow," *ACM Computing Surveys*, vol. 27(3), pp. 433–467, 1996.

[3] Y. Wei and L. Quan, "Region-based progressive stereo matching," *Proc. CVPR'04*, vol. 1, pp. 106–113, 2004.

[4] T. Cover and J. Thomas, *Elements of Information Theory*, New York: John Wiley & Sons, Inc., 1991.

[5] H. Jeffreys, "An invariant form for the prior probability in estimation problems," *Proc. of the Royal Soc. of London A*, vol. 186, pp. 453–454, 1946.

[6] Jason J. Corso, *Techniques for Vision-Based Human-Computer Interaction*, PhD dissertation, Johns Hopkins University, 2005.

[7] C. Canton-Ferrer, J.R. Casas, and M. Pardàs, "Towards a Bayesian approach to robust finding correspondences in multiple view geometry environments," *LNCS, Springer-Verlag*, vol. 3515, pp. 281–289, 2005.

[8] R.I. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*, Cambridge University Press, 2nd Ed., 2004.

[9] A. Ortega and K. Ramchandran, "Rate-distortion methods for image and video compression," *IEEE Signal Processing Magazine*, vol. 15, no. 6, pp. 23–50, Nov. 1998.

[10] H. Everett, "Generalized Lagrange multiplier method for solving problems of optimum allocation of resources," *Operations Research*, vol. 11, pp. 399–417, 1963.

[11] R.O. Duda, P.E. Hart, and D.G. Stork, *Pattern Classification*, New York: John Wiley & Sons, Inc., 2nd Ed., 2000.