Registration and interactive planar segmentation for stereo images of polyhedral scenes

Javier Flavio Vigueras and Mariano Rivera

Center for Mathematical Research
Centro de Investigacion en Matematicas (CIMAT)
Jalisco S/N, Colonia Valenciana
36240,Guanajuato, Gto, Mexico

Abstract

We introduce a two-step iterative segmentation and registration method to find coplanar surfaces among stereo images of a polyhedral environment. The novelties of this paper are: (i) to propose a user-defined initialization easing the image matching and segmentation, (ii) to incorporate color appearance and planar projection information into a Bayesian segmentation scheme, and (iii) to add consistency to the projective transformations related to the polyhedral structure of the scenes. The method utilizes an assisted Bayesian color segmentation scheme. The initial user-assisted segmentation is used to define search regions for planar homography image registration. The two reliable methods cooperate to obtain probabilities for coplanar regions with similar color information that are used to get a new segmentation by means of Quadratic Markov Measure Fields (QMMF). We search for the best regions by iterating both steps: registration and segmentation.

Key words: interactive computer vision, stereo, registration, segmentation, coplanarity, color

1 Introduction

Planar surfaces are often found in artificial manmade environments: outdoor scenes are commonly formed by polyhedral buildings (some examples are shown in Figure 1); indoor scenes contain floors, walls, desks, etc. Planes have a constrained representation and ease various computer vision tasks such as camera calibration [10,33], camera localization [13,30], robot navigation [24], and 3D reconstruction [7,36]. Plane-based algorithms are commonly stable

but they may become ill-conditioned when they are applied to wrong coplanar features, and therefore it is very important to know which image regions correspond to individual planes. By these reason, several works have been conducted on plane detection and segmentation.

1.1 Review of state-of-the-art segmentation approaches

A more general problem than the one issued in this paper is to estimate simultaneously the regions in the image corresponding to a given model (segmentation) and the set of parameters for each region model. In our specific case, each model corresponds to a planar surface.

If the model parameters are known, generic clustering algorithms as K-Means or Isodata have been used with relative success [12]. Other algorithms consider spatial interactions among pixel labels providing useful constraints on the problem: region merging [8], active contour [5] approaches, eigendecomposition [38] and variational methods [28]. Among these, Bayesian formulations [16,18] have been successfully used for finding the solution to the segmentation problem. In this framework, the solution is computed by Maximizing the A Posterior probability distribution (MAP estimator).

In the general case, the model parameters are not known and some of these methods are extended using two-step procedures, following that an initial estimate is given and then the method: (1) estimates the model parameters given the segmentation, (2) estimates the label map (segmentation) given the model parameters, iterating these two steps until convergence [6,20,35].

However, the MAP estimator for the label field requires the solution of a combinatorial optimization problem. Graph-Cuts based algorithms [6] can be used for computing the exact MAP estimator in the case of binary segmentation or an approximation for problems with more than two classes, but make the two-step algorithm prone to be trapped in local minima.

A better strategy is to compute, instead of binary label variables, the probability that the observed data at each pixel is generated by a particular model (i.e., the posterior marginal distributions). Posterior marginal probabilities can be estimated with Markov Chain Monte Carlo (MCMC) methods [16], by Mean Field (MF) approximations [35], or using a Gauss-Markov Measure Field (GMMF) model [19]. Nevertheless, MCMC approaches may be computationally expensive and the other two methods only guarantee convergence to a local maximum of the posterior distribution: if the procedure starts from a *wrong* initial set of parameters, the method may not converge to the global maximum [20]. Recently, the Quadratic Markov Measure Field (QMMF) approach was proposed [26], a computationally efficient method based on a

(a) (b) (c)

Fig. 1. Example of planar piecewise environments. Right images acquired by a stereo pair of a: (a) tower scene, (b) roof scene and (c) house scene.

Bayesian framework, improving GMMF model, since the posterior marginal probabilities are the global minimizer of a quadratic, linearly constrained energy function; hence, any standard linear algorithm will converge to the global minimum.

1.2 Previous works on plane segmentation

Unsupervised plane detection and segmentation are commonly solved using sparse image key-points (using structure from motion techniques [2,30,31]), disparity maps [14,34], optical flow approaches [41], triangular surfaces [22] or range images [37]. Nevertheless, many of these approaches require to perform a 3D reconstruction; other methods assume that the plane is mostly textured [2], that a single plane is dominant in the image, or the camera require a rough calibration [1,34], constraining the range of application of those methods.

Matching sparse features often fails in cases like the tower (Figure 1-a) and the roof (Figure 1-b) because both planes have mostly the same texture. **Although some heuristics (like RANSAC [9] or *a contrario* statistics [21]) have demonstrated to be successful on improving matching for a single surface having repeated structures (if a parametric model relating two views is given); their success is widely related with the ratio of inliers/outlier correspondences. However, when this heuristics are applied to distinguish between two or more planes, this ratio has a tendency to be very small and, as a result, the heuristics performance is noticeably poor.**

On the other hand, untextured surfaces (like the three walls in Figure 1-c) are also difficult to distinguish as different planes for optical flow techniques, disparity maps and feature-based methods, because image features are usually mismatched.

Image acquisition from a moving camera over a polyhedral scene imposes

known constraints on matching information for every couple of views and it is possible to extract planar segmentation from them without explicitly performing a 3D reconstruction [40]. Few approaches have tried to conduct segmentation on dense disparity models [32] but they often fail because several ambiguities arise on considering disparity information alone.

1.3 Interactivity

Several automatic solutions [4,6,15,27,39] have been developed for the planar segmentation problem from stereo views. Nevertheless, correct results are not reachable under some circumstances (described below in this paper), and only few automatic approaches may attain satisfactory results but in an unmanageable amount of time for some applications [17,29]. On the other hand, humans are capable to distinguish correctly distinct planar surfaces from a single image in a very short time. However, doing this task as a manual procedure could be unpractical. A completely user-assisted segmentation may be tedious if the boundaries of planar surfaces are not clearly distinguished into the image, when these boundaries have complex shapes, or when the interfaces are not friendly enough.

Computational complexity is not the only problem to cope. Automatic segmentation algorithms often have problems to make a decision on the number of region models to employ, thus, the number of planar surfaces observed in the images. Some approaches tend to over-segment the image, while others tend to merge different planes [17]. Interactive approaches are also very useful in this sense because allow the user to designate by himself/herself the number of planes in the images, or correct an over-segmented or under-segmented partition by assistance.

It often happens also in stereo that a plane is visible in one image but invisible in another view. Without an appropriate occlusion management, the matching process certainly may not succeed or may match to a wrong region in the second image; hence, the corresponding planar projective transformation (homography) may be wrong due to the occlusion. This is one of the major problems to cope with the automatic initialization processes. The interactive region selection has the advantage that the user may prevent such a problem with his/her previous knowledge of the scene.

Several computer vision methods seem to be very sensitive to the initial estimations required as input and, for many applications, choosing the adequate initial values may become an important or a tedious problem by itself. Interactive approaches have allowed to separate clearly the initialization stage from the automatic image processing, and improve the development of efficient computer vision tools due to the inclusion of knowledge given by human experts. **We have chosen an assisted strategy to initialize our algorithm, although automatic planar segmentation methods have been found in the literature: interactivity may help to reduce computational time and to correct wrong segmentations.**

In order to cope with the problems stated in 1.2, we propose a user-assisted segmentation method combining color information and motion matching of observed coplanar features in a two-image set. The novelty of this paper is to directly compute both homography parameters and dense region segmentation by means of a brief interaction with the user, instead of using unsupervised state-of-the-art methods that generally consider either case: (i) finding sparse coplanar points and then fitting a planar surface (implying occlusion and convexity problems), or (ii) computing general optical flow or disparity maps and later trying to fit a planar surface.

Our approach is based on a rough solution given into an initial stage by the user and an iterative two-step algorithm: (1) registering two views of the scene in order to find the corresponding planar homographies for every plane and (2) segmenting the image using a Bayesian segmentation approach. We refine the homography models using the new marginal a posteriori probabilities obtained from segmentation as a registration mask, and repeat the two-step procedure until convergence. As input for the planar segmentation step, model likelihoods are estimated by combining planar homographies fitting between the two views and their corresponding color information.

For the algorithm proposed in this paper, we have selected the Quadratic Markov Measure Field (QMMF) segmentation approach. QMMF performance was already demonstrated by numerical experiments that compare this approach [26] with other state-of-the-art algorithms, such as Graph-Cut [6], Random Walker, Gaussian Markov Measure Field [19] and Hidden Markov Measure Field methods [20].

1.5 Outline of the paper

The paper is organized as follows: our scheme for simultaneous segmentation and image registration for multi-planar environments is given in section 2. Next sections present the theoretical framework for parametric image registration (section 3), likelihood computation (section 4) and the Quadratic Markov Measure Fields (QMMF) approach for Bayesian segmentation (section 5). Section 6 presents a framework for computing homographies with polyhedral consistency. Section 7 exhibit results for some examples. Finally, a discussion on the method is summarized in section 8.

2 Overview

The particular problem issued in this paper may be formulated as follows: given some user-defined samples of coplanar regions taken from uncalibrated stereo views of a scene, we aim to estimate the corresponding planar projective transformations (homographies) and to extend the coplanar regions by segmenting the 2D stereo images. The vision system is modeled using the classical pin-hole camera model, which intrinsic parameters are supposed unknown and not necessarily constant.

Our approach has two different stages:

(1) **A user-assisted initialization step: The main objective of this initialization step is to include human knowledge about the regions in the image corresponding to planar surfaces. The human-machine interaction was designed in a very simple way: in one of both images to register, the user should either click once inside each planar region (and the system will automatically take a circular sample around the clicked position, see Figures 3-a and 8-a) or draw a complex detailed hull (see Figure 8-c), in order to define the sample region associated to a planar surface.**

(2) **When the region samples are already defined, the system searches automatically the planar projective transformation that match such a sample region with its corresponding region at the second image. An iterative registration-segmentation algorithm (described in Table 1) is performed. This two-step stage is iterative and completely automatic after the user has initialized the region samples. The algorithm is based in a region-growing approach and works independently to the initialization step.**

In our approach, the only required interaction is the user marking some sam-

ples of coplanar regions over one of the images at the procedure's beginning. Such a simple interaction provides very important information: the numbers of planes in the scene and small regions that undoubtedly belongs to each plane. Then, a Bayesian approach based on QMMF (see section 5) is applied in order to estimate the marginal probabilities based on color information around the user-defined sample regions. Once the coplanar regions were defined and the color probability fields were obtained, fully automated image registration and segmentation are iteratively done.

The core of our algorithm is an energy function relating the model regions and the respective homographies, an iterated minimization allows us:

(1) to compute the planar projective transformation (homography) coefficients corresponding to each identified region, and thus matching region intensities between both images;
(2) to compute new marginal probabilities that take into account spatial coherence. At the end of this stage, new *a posteriori* probabilities involving color and planar information are used to redefine registration masks for step 1.

The main steps of this algorithm are described in table 1.

3 Image registration

3.1 Parametric registration models

Camera model. In this paper, we consider that the camera model is the pinhole (perspective projection) model, which associates a point $\mathbf{X}$ in the scene to a point $\mathbf{x}$ in the image by $\tilde{\mathbf{x}} \sim \mathbf{P}\tilde{\mathbf{X}}$, where $\mathbf{P}$ is a 3×4 matrix called the *projection matrix*, $\tilde{\mathbf{x}}$ and $\tilde{\mathbf{X}}$ are 2D and 3D points, respectively, expressed in homogeneous coordinates: $\tilde{\mathbf{x}} = (\mathbf{x}, 1)^T$ and $\tilde{\mathbf{X}} = (\mathbf{X}, 1)^T$. $a \sim b$ means that a and b are equivalent up to a scale factor, thus, there exists a non-zero scalar λ such that $a = \lambda b$.

Planar projective transformations (homographies). Let us now constrain $\mathbf{X}$ to lie on a plane π (see Figure 2). If $\mathbf{x}_i = (x_i, y_i, 1)$ and $\mathbf{x'}_i = (x'_i, y'_i, 1)$ are two projections of the same 3D point $\mathbf{X}$ on the space for the first and the second image respectively, the projections are related by $\mathbf{x'}_i \sim \mathbf{H}\mathbf{x}_i$, then, the transformation modeling the 2D movement of coplanar points under perspective projection is given by a 3×3 homography matrix $\mathbf{H} = \{h_{ij} | i, j = 1, 2, 3\}$.

For each observed plane, a homography is implicitly related to the cameras

(1) Acquire two images of a planar piecewise environment.

(2) At one image, the user selects samples of the coplanar regions (see Figures 3-a, 8-a and 8-c) to be used in the registration process.

(3) Estimate the marginal probabilities (by QMMF, section 5) corresponding to the color information at user-defined regions (section 4.1).

(4) Step loop computation of the registration parameters (homography) and the segmented regions:

(a) Registration from coarse scale to finer scales:
- Establish the energy function for every defined region (Equation (2) for independent homographies and Equation (13) for the consistent approach).
- Compute the coefficients of each planar homography by minimizing the corresponding energy function.
- Update parameters for the multi-scale scheme.

(b) Segmentation:
- Compute likelihoods for planar homographies (section 4.2).
- Compute the composed likelihoods by the product of planar likelihoods and color probabilities (section 4.3).
- Estimate the composed marginal probabilities by QMMF Bayesian segmentation (section 5).
- For each model, select the biggest connected region with marginal probabilities greater than 0.5 (using a fill method). These regions will be used at the registration step instead of the initial color probability samples.

(5) (Optional) Refine region boundaries by using intersections between planes (section 6.3).

Table 1
Overview of the method.

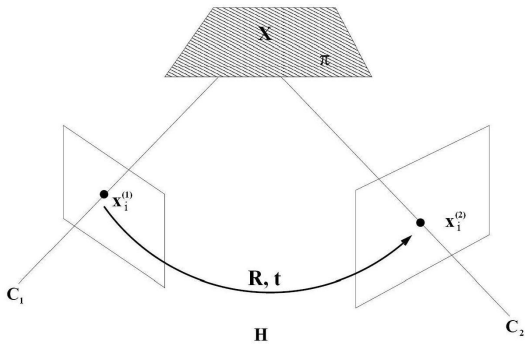


Fig. 2. Homography induced by a plane.

relative position, the plane position and the cameras projective properties [11]. Nevertheless, the coefficients of **H** may be directly estimated from image

correspondences and it is not necessary to know explicitly the camera and plane parameters.

Affine model. Another very common transformation used for image registration is the affine model ($h_{31} = h_{32} = 0, h_{33} = 1$), although this model is not always capable to describe the correspondences due to perspective projections. The affine model may be used when the image planes for both cameras lie on the same plane; it also represents a good approximation if both cameras have large focal lengths and the distance between the scene and the cameras is very large.

3.2 Energy function for parametric registration

Standard intensity-based approaches are founded on the luminance (chromance) constancy condition given as $\mathbf{I}_2(\mathbf{T}(\mathbf{x}, \Theta)) = \mathbf{I}_1(\mathbf{x}) + \eta(x)$ where $\mathbf{I}_1$ and $\mathbf{I}_2$ are the gray-level (color) intensities for the first and the second views, respectively, at a given image location $\mathbf{x}$ in the image lattice $\mathcal{L}$, $\mathbf{T}$ is the parametric transformation (with parameters Θ) describing the displacement of a 2D point in the first image to its corresponding projection at the second image, and η is identically distributed image noise for gray-level intensities.

Energy-based registration methods are based on the principle that two image regions are related by a given transformation T and that it is possible to compose an energy functional E in terms of the parameters Θ and the image data. These approaches rephrase the registration problem into an optimization problem where the expected set of parameters $\hat{\Theta}$ corresponds to the global minimum of the function E. Our registration is based on a cost function written as a weighted sum of squared differences of the above intensities given as follows:

$$E(\Theta) = \sum_{\mathbf{x} \in \mathcal{L}} \left[w(x) \left\| I_2(\mathbf{T}(\mathbf{x}, \Theta)) - I_1(\mathbf{x}) \right\|^2 \right] , \tag{1}$$

where $w(x) \in [0, 1]$ acts as a membership variable for each x in $\mathcal{L}$.

3.3 Retrieving the homographies

To retrieve the homographies, for each point $\mathbf{x}_i$ in the first image, a corresponding point $\mathbf{x}'_i$ is needed in the other image. **The energy-based approaches founded on luminance (or color) constancy do not need explicitly such image correspondences. Coordinates $\{\mathbf{x}'_i\}$ are expressed in terms of points $\{\mathbf{x}_i\}$ and the transformation parameters Θ: $\mathbf{x}'_i =$**

$\mathbf{T}(\mathbf{x}_i, \Theta)$. These approaches assume that the energy function (1) decreases when reprojection errors $\|I_2(\mathbf{T}(\mathbf{x}_i, \Theta)) - I_1(\mathbf{x}_i)\|$ are close to zero as a consequence of a correct registration, and rises for wrong transformation parameters.

By combining the cost function given by Equation (1) and the parametric transformation models (section 3.1), we get:

$$E(\mathbf{H}_k) = \sum_{\mathbf{x}\in\mathcal{L}} \left[w_k(x) \left\| I_2(\mathbf{x}') - I_1(\mathbf{x}) \right\|^2 \right] \tag{2}$$

where

$$\mathbf{x}' = z(\mathbf{H}_k, x, y) \begin{pmatrix} h_{11}^{(k)} x + h_{12}^{(k)} y + h_{13}^{(k)} \\ h_{21}^{(k)} x + h_{22}^{(k)} y + h_{23}^{(k)} \end{pmatrix}$$

N is the number of calibration planes, $\mathbf{H}_k$ is the transformation matrix for model k, $w_k(x) \in [0,1]$ acts as a membership variable for the $k-th$ planar region (the computation of such memberships is addressed in Section 5.2). The perspective division quotient is $z(\mathbf{H}_k, x, y) = \left(h_{31}^{(k)} x + h_{32}^{(k)} y + h_{33}^{(k)} \right)^{-1}$ for the homography model and $z(\mathbf{H}_k, x, y) = 1$ for the affine model.

The homography induced by each plane is computed by minimizing each energy function $E(\mathbf{H}_k)$. A non-linear optimization method (e.g. Levenberg-Marquardt) should be used to minimize these functions. Cost functions (2) will be useful for defining the QMMF energy in Equation (9).

4 Likelihood estimation

4.1 Likelihood associated to color information

Let I be an image such that $I(x) \in t_1 \cup t_2 \cup \cdots \cup t_T$, where $\{t_1, t_2, \ldots, t_T\}$ are the pixel values (represented by T disjoint subsets). **In our examples, pixel values are RGB-color for *sawtooth* and *venus* images in Figure 3 and for Figures 1 and 8, and gray-scale values for the cheerios pair in Figure 3.** Then, the density distribution for the color (gray-scale) classes are empirically estimated by using a histogram technique. That is, by smoothing the histograms such that g_{ki} is the number of hand labeled pixels with value t_i for the k-th class, then the normalized histogram is computed with:

$$\hat{g}_{ki} = \frac{g_{ki}}{\sum_{j=1}^{T} g_{kj}} \quad . \tag{3}$$

The likelihood function associated to color for each class k is then computed with:

$$v_k^C(x) = \frac{\hat{g}_{ki} + \epsilon}{\sum_{j=1}^{N}(\hat{g}_{ji} + \epsilon)} \quad \text{only if} \quad I(x) \in t_i, \tag{4}$$

where ϵ is a small positive constant (e.g. 1×10^{-8}) introduced to avoid the undefined computation of v_k^C when $\sum_{j=1}^{N} \hat{g}_{ji}$ becomes zero. This sum equals zero if a given color (or gray level) is not present in any of the hand labeled regions; then, ϵ stabilizes the likelihoods forcing them to follow a discrete uniform distribution.

4.2 Likelihood associated to coplanarity

In order to compute the likelihood of a pixel image associated to a homography matching model between two images, it is necessary to establish a probabilistic measure to compare two pixels information. The likelihood of a pixel associated to a homography transformation can be estimated as the probability that the given pixel is translated, at the second image, to another pixel with similar color information.

We assume that the probability distribution for noise $\eta^{(C)}$ on every color channel $A \in \{R, G, B\}$ is Gaussian with mean zero and standard deviation σ:

$$P(\eta^{(A)}(z)) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{z^2}{2\sigma^2}\right)$$

Then, taken into account only one-to-one pixel correspondences, the likelihood is stated as:

$$v_k^P(x) = \prod_{a \in \{R,G,B\}} P\left(\eta^{(a)}\left(I_1^{(a)}(x) - I_2^{(a)}\left(\mathbf{T}\left(x, \mathbf{H}_k\right)\right)\right)\right) \tag{5}$$

Likelihoods considering neighborhood texture information around x can be computed over a window of a given size (W_x, e.g. 3×3 pixels) by:

$$v_k^H(x) = \prod_{y \in W_x} v_k^P(y) \quad . \tag{6}$$

4.3 Composed likelihood: coplanarity and color

Trying to distinguish coplanar regions only through the planarity likelihood is not always a solvable problem, because a pixel projected by a homography

into the second image may fall into a region with a uniform color (e.g. the sky, untextured walls) and, in this case, the coplanarity score could be high even if the planar model is not correct. It means that the coplanarity likelihood is only informative at non-flat regions, mainly in high-gradient regions.

In order to propose a more informative criterion, we estimate a joint likelihood between planar and color information for each plane, such that the composed likelihood is computed as the product of Equation (4) and Equation (6):

$$v_k(x) = v_k^C(x) v_k^H(x) \tag{7}$$

5 Quadratic Markov Measure Fields (QMMF)

5.1 Computation of the Measure Probability Field

Given a set of likelihoods containing information about plane and color similarity to a given model, the goal is to find a probability field $\mathbf{p}$ indicating which model is supported for every pixel in the image. In particular probability $p_k(x)$ for pixel x will be the highest among $\mathbf{p} = \{p_i(x) : i = 1, \cdots, N\}$ if $x \in R_k$, where R_k is the region in the image that corresponds to model k, and N is the number of models. As $\mathbf{p}$ is a probability measure field, it has to satisfy the constraints:

$$\sum_{k=1}^{N} p_k(x) = 1 \qquad and \qquad p_k(x) \geq 0 \quad \forall k, x \tag{8}$$

According to the QMMF framework proposed in [26], the optimal estimator for the probability measure field is the minimum of the function:

$$U(p) = \sum_{x \in L} \left\{ \sum_{k=1}^{N} \left[-\log\left(v_k(x)\right) - \mu \right] p_k^2(x) + \lambda \sum_{y \in \mathcal{N}_x} \left\| p(x) - p(y) \right\|^2 \right\} \tag{9}$$

subject to the constraints (8). In Equation (9), $v(x) = (v_1(x), \cdots, v_N(x))$ is the likelihood vector with:

$$v_k(x) = P\left(I(x) | x \in R_k, \theta_k\right)$$

θ_k are the parameters for model k, λ is a non-negative parameter controlling spatial coherence between all the neighbors pixels $\langle x, y \rangle$, and the parameter μ is added to control the entropy of the probability field.

Gauss-Seidel method may be used for the minimization of the function given by Equation (9) using the iterative rule [26]:

$$p_k(x) = \frac{n_k(x)}{m_k(x)} + \frac{1 - \sum_{i=1}^{N} \frac{n_i(x)}{m_i(x)}}{m_k(x) \sum_{i=1}^{N} \frac{1}{m_i(x)}}$$

where:

$$\begin{aligned} n_k(x) &= \lambda \sum_{y \in \mathcal{N}_x} p_k(y) \\ m_k(x) &= -\log v_k(x) - \mu + \lambda \#(\mathcal{N}_x) \end{aligned}$$

To satisfy the constraints (8), all negative values should be set to zero and renormalization must be done if needed.

Relation with HMMF. Hidden Markov Measure Fields is a Bayesian segmentation scheme strongly related to QMMF. In [26], it is shown that for vectors $\mathbf{p}(x)$ with low entropy, the optimal HMMF estimator is approximated by the QMMF solution.

Binary QMMF. The general QMMF approach should satisfy the constraints (8). Nevertheless, the sum of probabilities is one only if every pixel at the image corresponds to an observation model. In the planar case, it means that all the observed surfaces should be planar (or approximated by planes) and that there would be a plane-color model corresponding to it. In other case, some pixels may be misclassified.

In the cases where just a few planes were required or the image is not completely formed by plane surfaces, we use a QMMF approach for binary classification [25]. Instead of computing simultaneously the probability field for N models, for each given model, we segment the image in two regions: one belonging to this model and another one for pixels that do not correspond. This approach will conduct N binary QMMF segmentation procedures and is faster than the general QMMF method:

$$\begin{aligned} U'(p_k) = \sum_{x \in L} \{ & [-\log(\hat{v}_k(x)) - \mu] \, p_k^2(x) \\ & + [-\log(\hat{u}_k(x)) - \mu] \, (1 - p_k(x))^2 \\ & + \lambda \sum_{y \in N_x} (p_k(x) - p_k(y))^2 \} \end{aligned} \tag{10}$$

where the input likelihoods for each binary segmentation are computed as

follows:

$$\begin{aligned} \hat{v}_k(x) &= \frac{v_k(x)+\epsilon}{\sum_{j=1}^{N}(v_j(x)+\epsilon)} \\ \hat{u}_k(x) &= \max\left\{\hat{v}_i(x) | i \neq k\right\} \end{aligned} \qquad \forall k = 1, \cdots, N \ . \tag{11}$$

5.2 Model parameter computation

The QMMF model allows to estimate the likelihood function (11) parameters [26]. To make explicit the parameter dependency of the QMMF energy functional (9), the first term is expressed as:

$$\sum_{x \in L} \sum_{k=1}^{N} \left\{ p_k^2(x) \left[\sum_{y \in W_x} \left\| I_2(\mathbf{T}(\mathbf{y}, \Theta)) - I_1(\mathbf{y}) \right\|^2 - \log\left(v_k^C(x)\right) - \mu \right] \right\}$$

by neglecting independent terms in Θ:

$$U(\Theta) = \sum_{x \in L} \sum_{k=1}^{N} \left\{ w_k(\mathbf{x}) \left\| I_2(\mathbf{T}(\mathbf{x}, \Theta)) - I_1(\mathbf{x}) \right\|^2 \right\}$$

where $w_k(\mathbf{x}) = \sum_{y \in W_x} p_k^2(y)$. This energy (12) is similar to $E(\mathbf{H}_k)$ in (2).

6 Our approach of projective representation for polyhedral structures

6.1 Dependency between homographies

In section 3, we have proposed a registration method based on the minimization of a function expressed in terms of the coefficients of each homography transformation $\mathbf{H}_k$. Nevertheless, like most of the multiplanar techniques found in the literature [1,2,17,24,32,34,40,41], this basic method does not take into account yet the fact that the set of homographies associated to different planes observed in both views are not independent and the results may be not consistent with the epipolar geometry of the camera array.

Each planar homography is associated to the cameras intrinsic parameters, relative motion between both views, and plane equations, as follows [11]:

$$\mathbf{H}_k \sim \mathbf{K}_2 \left(\mathbf{R} - \mathbf{t}\mathbf{v}_k^T \right) \mathbf{K}_1^{-1}$$

where $\mathbf{K}_1$ and $\mathbf{K}_2$ are the intrinsic parameter matrices for the first and the second cameras, respectively; $\mathbf{R}$ and $\mathbf{t}$ are the relative rotation and the relative translation between both views, resp.; and $\mathbf{v}_k$ is a vector associated to the $k-th$ plane, such that the plane equation is $\mathbf{v}_k \cdot \mathbf{x} + 1 = 0$. The motion and intrinsic parameters are common for all the possible planar homographies and the only different terms between these homographies are related with vectors $\mathbf{v}_k$.

Although there exist techniques to recover the whole set of parameters (intrinsic, motion and plane equations), these methods require a large number of either planes or views of the scene in order to obtain stable results. Nevertheless, self-calibration and structure from motion are not necessary to establish coherence for the homography set. In Ref. [3], the authors proposed a framework relating two planar homographies associated to a given epipolar geometry.

By using a similar approach, but avoiding the fundamental matrix computation and the stability problems associated to its estimation, we define a set of matrices $\mathbf{M}_{ij}$ such that $\mathbf{M}_{ij} = \mathbf{H}_i^{-1}\mathbf{H}_j$ for every couple of observed planes. Using the Sherman-Morrison formula, we obtain $\mathbf{M}_{ij} = \mathbf{I} + \mathbf{e}\mathbf{s}_{ij}^T$, where $\mathbf{s}_{ij} = \mathbf{K}_1^{-T}(\mathbf{v}_i - \mathbf{v}_j)$ and $\mathbf{e} = \mathbf{K}_1\mathbf{R}^{-1}\mathbf{t}$ is the right epipole of the given two-view geometry [11]. Consequently, we may express every planar homography in terms of a reference homography $\mathbf{H}_{ref}$, the epipole $\mathbf{e}$ and a 3D vector $\mathbf{s}_{ref,j}$ associated to the relative position between the reference plane and plane j:

$$\mathbf{H}_j \sim \mathbf{H}_{ref}\mathbf{M}_{ref,j} = \mathbf{H}_{ref}\left(\mathbf{I} + \mathbf{e}\mathbf{s}_{ref,j}^T\right) . \tag{12}$$

The reference homography can be any plane homography and the consistent homography estimation combines every independent function optimization (Equation 2) into the minimization of a unique function:

$$E\left(\mathbf{H}_{ref}, \mathbf{e}, \mathbf{s}_{ref,1}, \cdots, \mathbf{s}_{ref,N}\right) = \sum_{k=1}^{N}\sum_{\mathbf{x}\in\mathcal{L}}\left[w_k(x)\left\|I_2\left(\mathbf{x'}_k\right) - I_1\left(\mathbf{x}\right)\right\|^2\right] \tag{13}$$

where $\mathbf{x'}_k \sim \mathbf{H}_{ref}\left(\mathbf{I} + \mathbf{e}\mathbf{s}_{ref,k}^T\right)\mathbf{x} = \mathbf{H}_{ref}\left(\mathbf{x} + \left(\mathbf{s}_{ref,k} \cdot \mathbf{x}\right)\mathbf{e}\right)$. The minimization step can be conducted by means of the Levenberg-Marquardt method that seems to be very stable for these functions.

Our method has two main advantages respect to the approach presented in Ref. [3]: (i) it is not necessary to explicitly estimate the epipolar geometry (fundamental matrix) for the camera array, which may be very unstable for some scene structures or camera movements; (ii) the parameterization does not depend on the finite or infinite nature of the epipoles, that actually changes the mathematical form of the homography dependency.

6.2 Intersections between planes

Once the homographies have been consistently computed as shown in the previous section, intersections between planes can be computed in a very easy way: intersection between planes i and j implies that a given point $\mathbf{x}$ belonging to their intersection at the first image satisfies $\mathbf{H}_i\mathbf{x} \sim \mathbf{H}_j\mathbf{x}$; thus $\mathbf{x} \sim \mathbf{H}_i^{-1}\mathbf{H}_j\mathbf{x}$, being $\mathbf{x}$ an eigenvector of $\mathbf{M}_{ij}$. The eigenvectors of $\mathbf{M}_{ij}$ are:

- the epipole $\mathbf{e}$, which is common for all the possible couples of planes: $\left(\mathbf{I} + \mathbf{e}\mathbf{s}_{ij}^T\right)\mathbf{e} = (1 + \mathbf{s}_{ij} \cdot \mathbf{e})\,\mathbf{e}$
- every vector $\mathbf{a}$ orthogonal to $\mathbf{s}_{ij}$: $\left(\mathbf{I} + \mathbf{e}\mathbf{s}_{ij}^T\right)\mathbf{a} = \mathbf{a} + (\mathbf{s}_{ij} \cdot \mathbf{a})\,\mathbf{e} = \mathbf{a}$

The intersection between planes i and j is the set of all such vectors $\mathbf{a}$. Furthermore, by orthogonality we know that $\mathbf{s}_{ij} \cdot \mathbf{a} = 0$, then the intersection (line) equation is straight given by the vector $\mathbf{s}_{ij}$.

Our approach does not require the explicit computation of the eigenvectors of matrix $\mathbf{M}_{ij}$, which may be unstable for several cases. Using the parameters obtained from the minimization of Equation (13):

- the intersection between the reference plane and a given plane i, observed at the first image, is directly given by $\mathbf{s}_{ref,i} \cdot \mathbf{x} = 0$;
- the intersection between any other planes i and j are given by the eigenvectors of $\mathbf{M}_{ij}$:

$$\begin{aligned}\mathbf{M}_{ij} = \mathbf{M}_{ref,i}^{-1}\mathbf{M}_{ref,j} &= \left(\mathbf{I} + \mathbf{e}\mathbf{s}_{ref,i}^T\right)^{-1}\left(\mathbf{I} + \mathbf{e}\mathbf{s}_{ref,j}^T\right) \\ &= \mathbf{I} + \frac{\mathbf{e}\left(\mathbf{s}_{ref,j} - \mathbf{s}_{ref,i}\right)^T}{1 + \mathbf{s}_{ref,i} \cdot \mathbf{e}}\end{aligned}$$

  Thus, the intersection line is defined by Equation $(\mathbf{s}_{ref,j} - \mathbf{s}_{ref,i}) \cdot \mathbf{x} = 0$.

6.3 Improving boundary detection

In order to get a better segmentation for scenes containing polyhedral structures, we also propose an improvement to the original QMMF method. The approach consists on avoiding probability propagation through the plane intersections and delaying the propagation if there is not enough information to decide to which region a pixel belongs. We reach this goal by changing the spatial coherence term in Equations (9) and (10) to:

$$\lambda \sum_{\langle x,y \rangle} z(x,y)\,\|p(x) - p(y)\|^2$$

where $\langle x, y\rangle$ represents two neighbor pixels x and y (left, right, up and down neighbors), and $z(x, y)$ is a function controlling the probability propagation between these two pixel positions.

The approach is divided into two stages:

(1) For every iteration at the segmentation step, $z(x, y)$ is computed by $z(x, y) = \hat{\mathbf{v}}(x) \cdot \hat{\mathbf{v}}(y)$, where $\hat{\mathbf{v}}^T(x) = (\hat{v}_1(x), \cdots, \hat{v}_N(x))$ is the vector of normalized likelihoods at pixel x.
- Normalization means that $\sum_{k=1}^{N} \hat{v}_k(x) = 1$, thus the maximum value for $z(x, y)$ is reached only when $max\{\hat{v}_k(x)\} = max\{\hat{v}_k(y)\} = 1$; it means that the propagation will be bigger when both likelihoods associated to the same segmentation model are close to 1.
- If two or more segmentation models have similar likelihoods for neighbor pixels x and y, $z(x, y)$ will reduce the propagation between them. For example, if $\hat{v}_k^T(x) \approx \hat{v}_k^T(y) \approx \left(\frac{1}{2}, \frac{1}{2}, 0, \cdots, 0\right)$, then $z(x, y) \approx 2(\frac{1}{2})^2 = \frac{1}{2}$.
- If likelihood vectors indicate that pixel x is assigned to a given model and pixel y is assigned to a different one, $z(x, y)$ will approximate to zero, retarding the propagation between two different regions.

(2) At a final stage, we use the boundary information as given in Section 6.2 in order to refine the segmentation obtained at the iterative process. We assign a very small value ϵ to $z(x, y)$ when the pixels x and y are separated by the intersection of planes i and j, only if x and y were assigned to one of these planes (both assigned to the same plane, or x assigned to a plane, and y assigned to the other one).

7 Experimental results

In this section, we present the results of segmenting stereo-pair images with the proposed method. The images present on classical benchmarks are mostly textured, and then most of the methods are not allowed to cope with problems like segmenting: single-color surfaces, different coplanar models with similar texture, or single surfaces containing several textures or colors. In our experiments, we selected stereo pair views showing special difficulties (like (i) a single plane containing different textures or colors, or (ii) different planes having the same texture, e.g the roof) that do not appear in classical benchmarks.

In our implementation, all the routines were written in Turbo C++ 6.0 based on the numerical library in "Numerical Recipes". We opted for using the Levenberg-Marquardt method [23] to minimize the Equation (2). **We propose each $\mathbf{H}_k$ equal to the identity to initialize the optimization process, although the energy function optimization seems to be very sensitive to the initial set of parameters.** In order to avoid converg-

ing to local minima, we propose to use a multi-scale approach on matching both images, as it has already been used in optical flow estimation. Firstly, the image is sub-sampled four levels (each dividing the image size by two) and a homography is estimated for a coarse scale; the coarse scale solution should be then used as initial parameters for solving the registration problem at a finer scale.

7.1 Results on the Middlebury database

Firstly, we tested our approach on a set of three stereo pairs provided at the Middlebury dataset [29] (available online at http://www.middlebury.edu/stereo/).

For these three stereo pairs, the user has defined manually some sample regions (color circles at Figures 3-a) lying in some planar surfaces. The interface is very easy: the user click on a given point and a circular region is drawn around the point with a fixed diameter (25 pixels in these experiments). The standard deviation used to compute the registration likelihood (Equation 5) is fixed to $\sigma = 10$, and the binary QMMF parameters are $\lambda = 10$ and $\mu = 4$.

Stereo pair	Resolution	Data space	# of regions	# of iterations	Step 3, color likelihood (sec)	Step 4, iterative process (sec)	Total time (sec)
cheerios	320×240	gray	5	10	43	686	729
sawtooth	434×380	RGB	3	10	51	668	719
venus	434×383	RGB	5	10	89	1174	1263

Table 2
Time of computation for results on Figure 3-d.

The stereo pairs show several slanted planes with varying amounts of texture (even with no texture as in some regions of the *venus* scene). For these three pairs, we computed the input likelihoods using uniquely **coplanarity** information (no color information has been added). The results obtained after 10 iterations of our approach are shown in Figure 3-d and the times for the automatic stage are shown in Table 2.

Region boundaries are better defined for the Lin's technique, because it has included a term for boundary smoothness that favors the developing of rectilinear boundaries. Nevertheless, real acute-angled or non-straight boundaries are straightened by the method because of this term, as we can see for the *sawtooth* and *venus* examples (Figure 3-g): acute-angled ends are truncated.

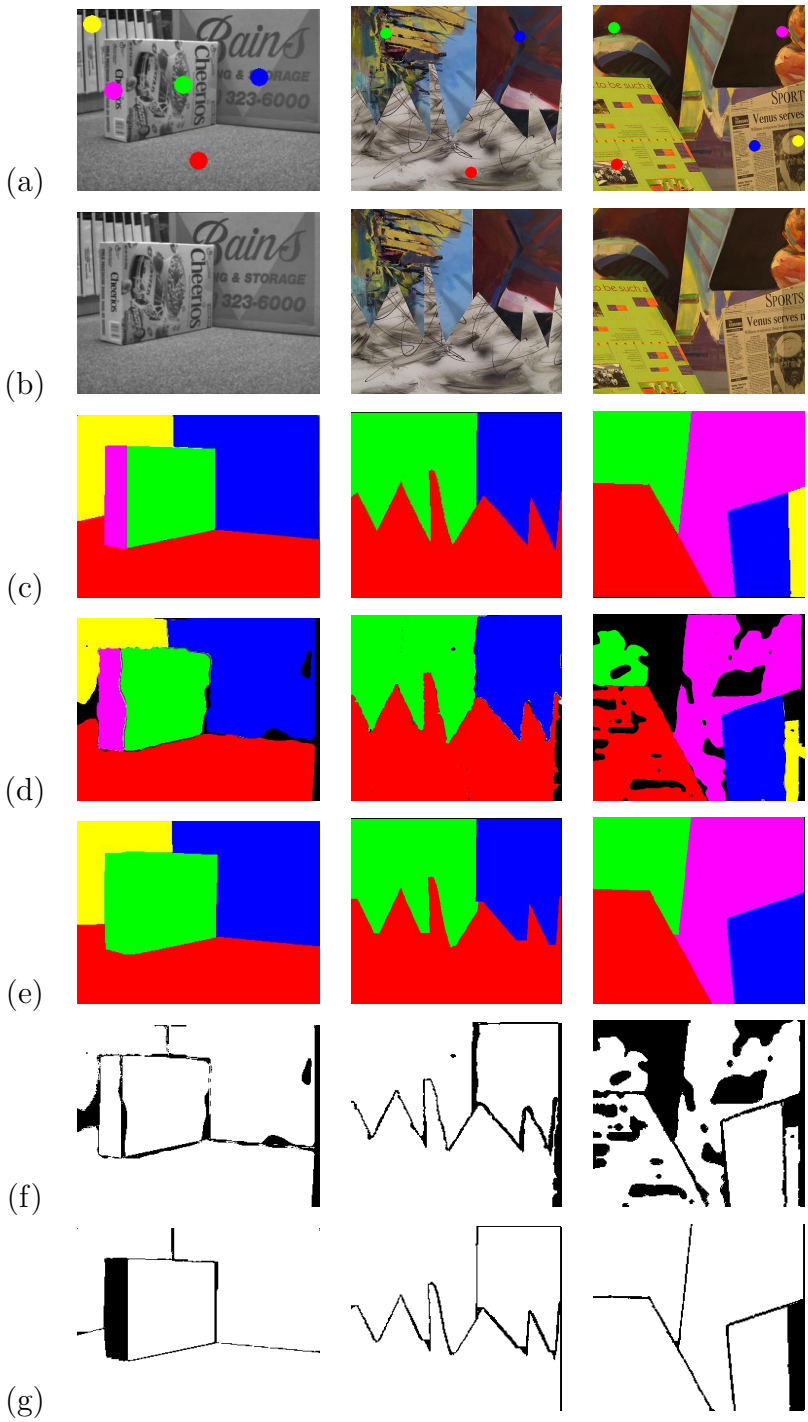
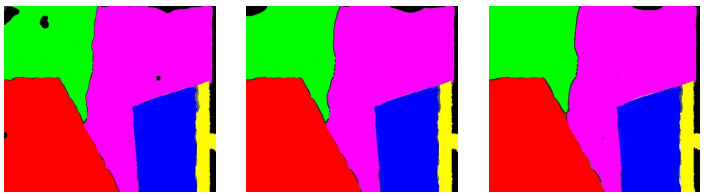


Fig. 3. Images of the Middlebury database: *cheerios*, *sawtooth* and *venus* (from left to right). (a) Right images and (b) user-defined sample regions at left images (circles with diameter = 25). Segmentation: (c) ground-truth, (d) after ten registration-segmentation iterations of our method, (e) for Lin's method. Difference images (mismatched pixels shown in black): (f) for our approach and (g) for Lin's approach.

Lin's method did not distinguish that pink and green regions correspond to different planes in the *cheerios* scene, as blue and yellow regions in the *venus*

pair are also different. The same problem has been found for both images in all the fully automatic methods [4,6,15,27] reported in [29]. Interactivity offers the advantage to indicate the procedure that these regions are different (as our method does, Figure 3-d) and the possibility to correct wrong tagging.



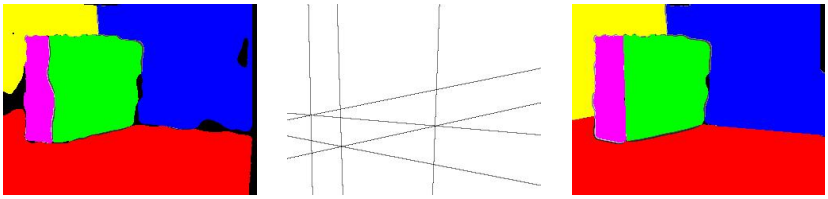
(a) (b) (c)

Fig. 4. Results for the *venus* scene using QMMF with $\mu = 0.4\lambda$ and different coherence parameters: (a) $\lambda = 10$, (b) $\lambda = 10^2$, and (c) $\lambda = 10^3$.

At the *venus* sequence, we see that the segmented planes for our method are hollow regions (Figure 3-d), mainly in image zones where there is no texture. Our method offers the option to force the pixels to be assigned to a given planar model, by changing the segmentation method from binary QMMF to several-models QMMF (Figure 4-a). **Observed gaps are due to occlusions or to not connected regions (because the algorithm just keeps the biggest connected regions with marginal probabilities greater than** 0.5**). Our interface gives to the user also the facility to control the coherence parameter λ and the entropy parameter μ interactively (allowing to repeat step 4(b) of the algorithm), if he/she considers that the segmentation is too granular or smoothed.** In Figure 4, we observed the results for different coherence values. We can see the advantage of an interactive segmentation in order to choose easily the best segmentation method (binary QMMF or multi-class QMMF) or to control the spatial coherence of the regions for a given stereo pair. Furthermore, it is not necessary to perform again the registration step when the user changes the segmentation strategy or the parameters, because the QMMF segmentation step is only related to the second stage of the iterative procedure.

Finally, we tested two different registration strategies: (i) considering that planar homographies are completely independent, (ii) taken into account homography dependency by using the projective framework presented in section 6. For the first registration approach, tests require several hundred of iterations, taking several hours before to converge to a stationary solution. On the other hand, homography consistency allows converging in very few iterations, as it is shown in Figures 3-d and 4 after 10 iterations. When convergence is reached, the solutions for both approaches are almost the same, such that differences are difficult to visualize; hence, the main advantage on considering homography consistency is on improving the computational time.

7.2 Refining the planar image segmentation



(a) (b) (c)

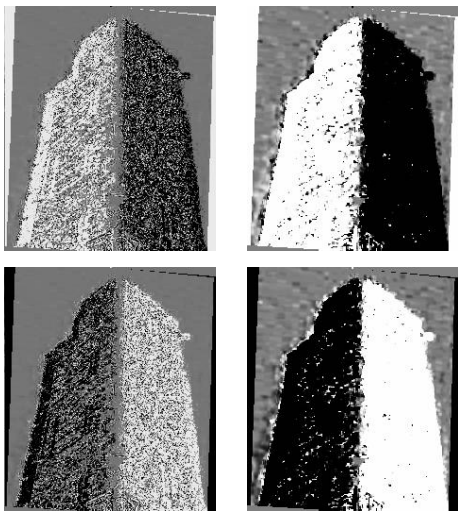
Fig. 5. Segmentation of the *cheerios* scene, (a) for ten iterations of our approach, (b) the map of plane intersections, (c) the image segmentation taking into account plane intersections.

In Figure 5-a, we appreciate that the obtained regions seem to be qualitatively similar to the ones stated in the ground truth. Nevertheless, the boundaries of the regions are not well defined. In order to improve the final segmentation, we add the refining step that takes into account the intersection between planes as it is indicated in Section 6.3. Figure 5-b shows the lines corresponding to plane intersections that are visible in the first image. At Figure 5-c, it is shown that the region growing is controlled in order to avoid propagation through the regions boundaries and it helps to improve the segmentation quality for polyhedral scenes.

7.3 Including color and texture to planar models

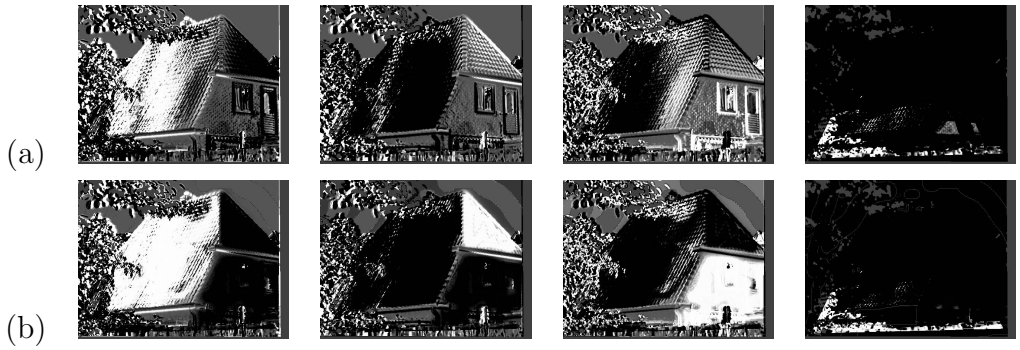
The tests were also conducted on another set of stereo views of real outdoor scenes. The experiment shown in Figure 6 is designed to demonstrate the importance of incorporating context information related with texture in the likelihood computation. Experiment shown in Figure 7 illustrates the importance of taking into account both color and coplanarity for computing likelihoods. In this section, we intend to show step by step how informative are coplanarity, texture and color likelihoods as inputs for the segmentation process.

Firstly, we compute the homographies corresponding to both observed planes at the tower scene by means of the procedure described at section 3. The normalized likelihoods obtained for these registration models, considering pixel-to-pixel similarities, are shown in Figure 6-a. White pixels represents highest similarities, while darker pixels correspond to lower planarity fitting to that model. Most of the pixels corresponding to the correct plane are shown in white; nevertheless, the regions are too granular and not connected. When using 3×3 regions (as it is shown in Figure 6-b), the coplanarity likelihoods are improved and well defined, such that they can easily be distinguished by very simple methods, like thresholding the images.



(a) (b)

Fig. 6. Normalized coplanarity likelihoods computed (a) pixel-to-pixel and (b) for 3×3 windows, for both planes of the tower, left and right.



(a)

(b)

Fig. 7. Likelihoods computed for (a) planar similarity and (b) coplanarity and color.

Nevertheless, normalized planarity likelihoods may be less informative if the regions are untextured or if there are different coplanar models with the similar texture. This is the case of the left and right sides of the roof (shown in Figure 1). In the images at Figure 7-a, only the pixels corresponding to high gradient points at the original images have high likelihood. Flat-color regions have almost the same likelihood for the different models, as it is observed for the sky pixels and the facade. Moreover, intersections between planes are not well defined, because the boundary regions may correspond to either plane at the junction. When color information is included, highest likelihoods distinguish better the complete planar surfaces and not only the high gradient elements, as it is shown in Figure 7-b.

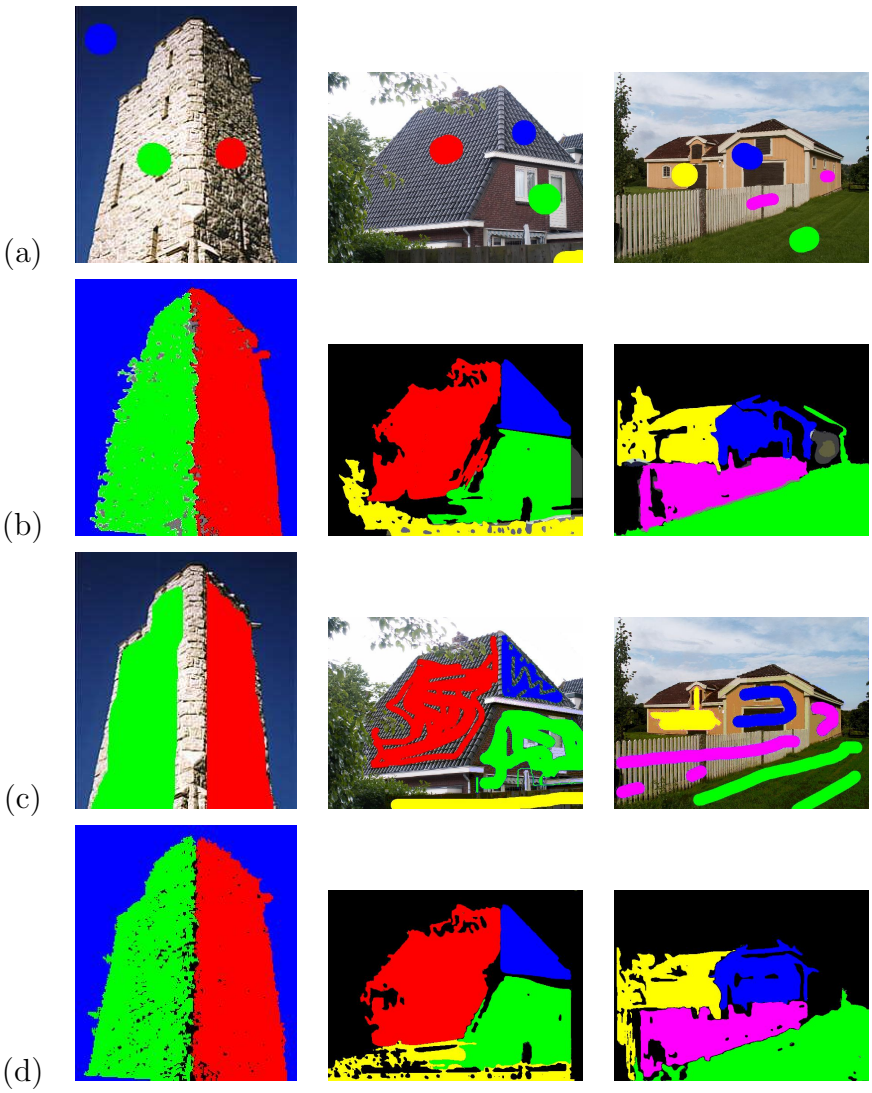


Fig. 8. Experiments for the: tower scene (first column), roof scene (second column) and house scene (third column). (a) Small user-defined sample regions at the left images and (b) their corresponding segmentations (after 10 binary QMMF iterations); (c) large user-defined sample regions and (d) the resulting segmentations (after just one binary QMMF iteration for each model). Note the segmented regions correspond to class membership larger than 0.5.

7.4 Results on outdoor images

At this step, maximum likelihood estimator (MLE) is good enough to distinguish most of the coplanar pixels with similar texture and color properties, but the detected regions are granular and not necessarily connected. In order to solve this problem, we apply the QMMF segmentation method (described in section 5). Once the a posteriori probabilities were obtained, we select the planar regions by selecting the pixels whose probability is greater or equal than 0.5 and then, the biggest connected region for each planar model.

If the results are not satisfactory at the end of the complete registration-segmentation process, the user can be capable to make corrections online of the segmentation, manually extend the sample regions, fill the gaps, etc. (e.g. compare user-defined samples in Figures 8-a and 8-c). This is one of the main advantages of interactivity.

For the final experiment, the user defined some image samples, shown in Figure 8-a, and the obtained results after ten binary QMMF iterations are shown in Figure 8-b. In the three sets of images, the planar regions were correctly extended, but there are still misclassifications if there are two concurrent planes or regions with similar texture or color. In Figure 8-c, the user extends the sample regions such that the results shown in Figure 8-d were obtained after only one binary QMMF iteration. Our results also show that the algorithm can distinguish planar surfaces from non-planar elements (e.g in the *roof* stereo pair, the tree leaves are correctly segmented out from the left region of the house roof).

8 Summary and conclusions

In this paper, we have proposed a method for segmenting and registering coplanar regions of a scene, based on the observation of a pair of stereo views. This process can be used in any environment containing planar structures. Its domain of application is wide because planar surfaces are quite common both at indoor and outdoor scenes. One can intrinsically segment a piecewise planar scene from two 2D images without performing neither camera calibration nor 3D reconstruction.

There are still limitations and further research to continue. We have chosen an assisted strategy to initialize our algorithm. Indeed, this approach depends on the user selection of sample regions and it may be sensitive to this choice. Although automatic planar segmentation methods have been found in the literature, interactivity may help to reduce computational time (from several hours [17] to some seconds in our method) and to correct wrong segmentations. However, we consider that this stage is independent of our method and that it does not demerit our proposal; contrarily, interactivity may help to extend planar regions that other methods cannot detect or separate. Our approach may accept any other initialization method and we are in actuality studying other approaches for this step.

We also have analyzed the information that can be extracted from color properties of the images and from move matching between two views of a planar surface. We propose to use an approach combining both kinds of information and to refine the marginal probability of finding both criteria for a given pla-

nar model, by means of an innovative Bayesian approach based on Markov Random Fields.

One of the future improvements of our method is that now registration has been conducted only from the first image to the second one, but no coherence from the registration in the opposite sense has been verified. This step should improve the performance of our method in some challenging regions such as discontinuities and occlusions. Another goal of our future work consists on obtaining metric information from the computed registration parameters, such as the epipole positions and the plane parameters, for a larger set of views, intending to auto-calibrate the cameras and to recover the planar piecewise structure of the scene.

Acknowledgments

This research was supported by CONACYT (grant 43743).

References

[1] C. Baillard and A. Zisserman, *"Automatic reconstruction of piecewise planar models from multiple views"*, in CVPR'99: Conference on Computer Vision and Pattern Recognition, pages 2559-2565. IEEE Computer Society, 1999. 3, 14

[2] A. Bartoli, *"A random sampling strategy for piecewise planar scene segmentation"*, Computer Vision and Image Understanding, 105(1):42-59, 2007. 3, 14

[3] A. Bartoli, P. Sturm and R. Horaud, *"A projective framework for structure and motion recovery from two views of a piecewise planar scene"*, Rapport de recherche no. 4070, INRIA Rhone-Alpes, 30 pages, 2000. 15

[4] S. Birchfield and C. Tomasi, *"Multiway cut for stereo and motion with slanted surfaces"*, ICCV, pages 489495, 1999. 4, 20

[5] A. Blake and M. Isard, *"Active Contours: The Application of Techniques from Graphics, Vision, Control Theory and Statistics to Visual Tracking of Shapes in Motion"*. Springer, 2000. 2

[6] Y. Boykov, O. Veksler, and R. Zabih, *"Fast approximate energy minimization via graph cuts"*, IEEE Trans. Pattern Anal. Machine Intell., vol. 23, no. 11, pp. 12221239, 2001. 2, 4, 5, 20

[7] R. Cipolla, T. Drummond and D. P. Robertson, *"Camera calibration from vanishing points in image of architectural scenes"*, in BMVC'99: Proceedings of the British Machine Vision Conference, volume 2: 382–391. Nottingham, UK. British Machine Vision Association, 1999. 1

[8] Y. Deng and B. S. Manjunath, *"Unsupervised segmentation of color-texture regions in images and video"*, IEEE Trans. Pattern Anal. Machine Intell., vol. 23, no. 8, pp. 800810, 2001. 2

[9] M. Fischler and R. Bolles, *"Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography"*, Communications of the ACM, vol. 24, pp. 381–385, 1981. 3

[10] P. Gurdjos and P. Sturm, *"Methods and geometry for plane-based self-calibration"*, in CVPR'03: Proceedings of the International Conference on Computer Vision and Pattern Recognition, volume 1, pages 491-496, Madison, US, 2003. IEEE. 1

[11] R. Hartley and A. Zisserman, *"Multiple view geometry in computer vision"*, Cambridge University Press, New York, NY, USA, 2000. 8, 14, 15

[12] A. K. Jain and R. C. Dubes, *"Algorithm for Clustering Data"*. Prentice Hall, Englewood Cliffs, NJ, 1998. 2

[13] K. Kanatani, *"Geometric computation for machine vision"*, Oxford University Press, Inc., New York, NY, USA, 1993. 1

[14] R. Koch, *"Surface segmentation and modeling of 3D polygonal objects from stereoscopic image pairs"*, in ICPR'96: Proceedings of the International Conference on Pattern Recognition, volume I, page 233, Washington, DC, USA, 1996. IEEE Computer Society. 3

[15] V. Kolmogorov and R. Zabih. *"Computing visual correspondence with occlusions using graph cuts"*, ICCV, volume II, pages 508515, 2001. 4, 20

[16] S. Z. Li, *"Markov Random Field Modeling in Image Analysis"*. Springer-Verlag, Tokyo, 2001. 2

[17] M. Lin, *"Surfaces with Occlusions from Layered Stereo"*, Ph.D. thesis, Stanford University, USA, 2003. 4, 14, 24

[18] J. Marroquin, S. Mitter, and T. Poggio, *"Probabilistic solution of ill-posed problems in computational vision"*, J. Am. Stat. Asoc., vol. 82, pp. 7689, Nov 1987. 2

[19] J. L. Marroquin, F. Velazco, M. Rivera, and M. Nakamura, *"Probabilistic solution of ill-posed problems in computational vision"*, IEEE Trans. Pattern Anal. Machine Intell., vol. 23, pp. 337348, 2001. 2, 5

[20] J. L. Marroquin, E. Arce, and S. Botello, *" Hidden Markov measure field models for image segmentation"*, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 25(11):1380-1387, Nov. 2003. 2, 5

[21] L. Moisan and B. Stival, *" A Probabilistic Criterion to Detect Rigid Point Matches Between Two Images and Estimate the Fundamental Matrix"*, IJCV, volume 07, issue 3, pages 201–218, 2004. 3

[22] A. Nakatsuji, Y. Sugaya, and K. Kanatani, *"Optimizing a triangular mesh for shape reconstruction from images"*, IEICE - Transactions on Information Systems, E88-D(10):2269- 2276, 2005. 3

[23] J. Nocedal, *"Numerical optimization"*, Springer Verlag, New York, NY, USA, 1999. 17

[24] K. Okada, S. Kagami, M. Inaba and H. Inoue, *"Plane segment finder: Algorithm, implementation and applications"*, in ICRA'01:International Conference on Robotics and Automation, pages 2120-2125, Seoul, Korea, 2001. IEEE. 1, 14

[25] M. Rivera and P. P. Mayorga, *"Quadratic Markovian probability fields for image binary segmentation"*, in Proc. ICCV'07, 2007. 13

[26] M. Rivera, O. Ocegueda, and J. L. Marroquin, *"Entropy controlled quadratic Markov measure field models for efficient image segmentation"*, IEEE Trans. Image Processing, 8(12):3047-3057, 2007. 2, 5, 12, 13, 14

[27] S. Roy. *"Stereo without epipolar lines: A maximum flow formulation"*, IJCV, 34(2/3):147161, 1999. 4, 20

[28] C. Samson, L. Blanc-Feraud, J. Zerubia, and G. Aubert, *"A variational model for image classification and restoration"*, IEEE Trans. Pattern Anal. Machine Intell., pp. 460472, 2000. 2

[29] D. Scharstein and R. Szeliski, *"A taxonomy and evaluation of dense two-frame stereo correspondence algorithm"*, International Journal of Computer Vision, 47:7-42, 2002. 4, 18, 20

[30] G. Simon, A. Fitzgibbon and A. Zisserman, *"Markerless tracking using planar structures in the scene"*, in ISAR'00: Proceedings of the International Symposium on Augmented Reality, pages 120–128. Munich, Germany, 2000. 1, 3

[31] D. Sinclair and A. Blake, *"Quantitative planar region detection"*, Int. Journal on Computer Vision, 18(1): 77-91, 1996. 3

[32] N. Thakoor, S. Jung and J. Gao, *"Real-time planar surface segmentation in disparity space"*, in CVPR'07: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pages 1-8. IEEE Computer Society, 2007. 4, 14

[33] B. Triggs, *"Autocalibration from planar scenes"*, in ECCV'98: Proceedings of the 5th European Conference on Computer Vision-Volume I, pages 89-105, London, UK, 1998. Springer-Verlag. 1

[34] E. Trucco, F. Isgro and F. Bracchi, *"Plane detection in disparity space"*, in VIE'03: International Conference on Visual Information Engineering, pages 73-76. IEEE, 2003. 3, 14

[35] Tsai, J. Zhang and A. Willsky, *"Expectation-Maximization Algorithms for Image Processing Using Multiscale Methods and Mean Field Theory, with Applications to Laser Radar Range Profiling and Segmentation"*, Optical Engineering, 40, 7, 1287-1301, 2001. 2

[36] C. Vestri and F. Devernay, *"Using robust methods for automatic extraction of buildings"*, in CVPR'01:Proceedings of the International Conference on Computer Vision and Pattern Recognition, volume 1: 133–138. 1

[37] X. Wang and H. Wang, *"Markov random field modeled range image segmentation"*, Pattern Recognition Lett., 25(3):367- 375, 2004. 3

[38] Y. Weiss, *"Segmentation using eigenvectors: A unifying view"*, in ICCV (2), 1999, pp. 975982. 2

[39] O. Woodford, P. Torr, I. Reid and A. Fitzgibbon, *"Global Stereo Reconstruction under Second Order Smoothness Priors"*, in CVPR'2008, pages 1–8. 4

[40] A. Y. Yang, S. Rao, A. Wagner and Y. Ma, *"Segmentation of a piece-wise planar scene from perspective images"*, in CVPR'05: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, volume 1, pages 154-161, Washington, DC, USA, 2005. IEEE Computer Society. 4, 14

[41] M. Zucchelli, J. Santos-Victor and H. I. Christensen, *"Multiple plane segmentation using optical flow"*, in BMVC'02: Proceedings of the British Machine Vision Conference, pages 313–322. Cardiff, UK, 2002. British Machine Vision Association. 3, 14