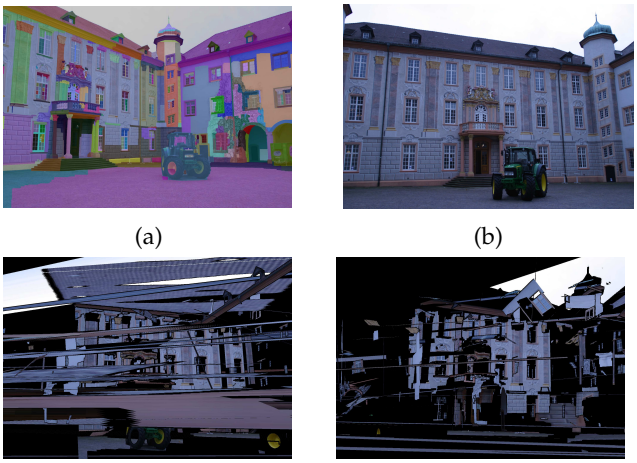
# Supplementary material: Piecewise-planar 3D approximation from wide-baseline stereo

C. Verleysen and C. De Vleeschouwer
ICTEAM institute, Université catholique de Louvain (UCL), Louvain-la-Neuve, Belgium
{cedric.verleysen, christophe.devleeschouwer}@uclouvain.be

## 1. Example of direct RANSAC-based plane fitting on segmented regions

As illustrated on the Castle sequence [33] in Figure 1, independent robust fitting of 3D planes (based on RANSAC [10]) over the regions is too sensitive to the strong presence of noisy and/or 3D outliers (points) in a dense point cloud obtained from only two wide-baseline cameras.



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

Figure 1: Fitting 3D planes on a dense point cloud generated from a pair of wide-baseline images is a challenging problem due to the implicit noisiness of the associated point cloud. (a) Segmentation of the left image; (b) Right image; (c) Left-to-right view projection, via the piecewise-planar 3D model of the scene obtained by using RANSAC [10] to fit a plane to the 3D points associated to each region; (d) Left-to-right projection, when the 3D planar model is derived from the 3D points corresponding to the most accurate and unambiguous[6] 2D matches. Even in this case, RANSAC is insufficient to approximate the 3D model of the scene.

[6]The notions of accuracy and ambiguity of a matching are defined in Section 4 of the paper.

## 2. Assessment of the plane quality metric

The quality metric $q(\pi_m)$ of a candidate plane $\pi_m$, which is the boundless version of the randomly selected triangular patch $[\pi_m]$, measures the fraction of 3D points $\mathbf{X}_j \in \{\mathbf{\Delta}_m \cup \mathbf{\Delta}'_m\}$ whose distance to the plane $\pi_m$is smaller than a predefined threshold $T_d \in \mathbb{R}^+$ (see Equation (3)). To illustrate the relevance of this metric, Figure 2 presents the distribution of the orthogonal distance $d(\pi_m, \mathbf{X}_j)$ between the 3D points $\mathbf{X}_j \in \{\mathbf{\Delta}_m \cup \mathbf{\Delta}'_m\}$ and the plane $\pi_m$, for two kinds of planes $\pi_m$: the 3D planes approximating correctly the ground-truth 3D surface (in green), and the ones that are far away from this ground-truth (in red).

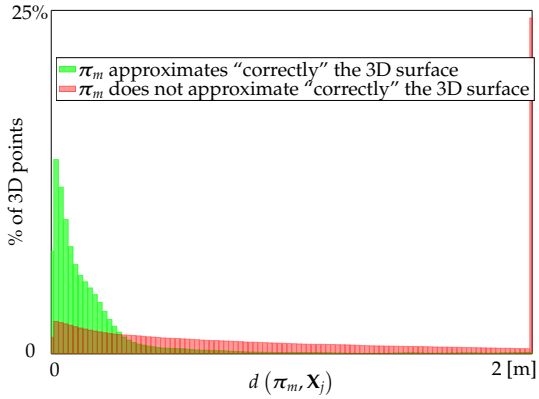



Figure 2: The quality $q(\pi_m)$ of a plane candidate $\pi_m$, associated to a randomly selected triangular patch $[\pi_m]$, is quantified based on the fraction of 3D points $\mathbf{X}_j \in \{\mathbf{\Delta}_m \cup \mathbf{\Delta}'_m\}$ that are close to this 3D plane. We validate the relevance of this measure by showing that there are much more 3D points $\mathbf{X}_j$ close to the "ground-truth" (definition detailed in the text here below) 3D planes (green histogram) than around any arbitrary other one (red histogram).

In Figure 2, a plane candidate $\pi_m$ is considered to approximate a ground-truth plane $\pi_i^\star$ when its orientation is close to the one of the ground-truth plane and

its distance to an arbitrary 3D point (*e.g.*, the optical center $\mathbf{C}$ of the source camera) is similar to the one of the ground-truth plane. In practice, $\boldsymbol{\pi}_m$ approximates "correctly" the ground-truth when there exists (at least) one ground-truth plane $\boldsymbol{\pi}_i^\star$ such that:

$$\Theta_{\boldsymbol{\pi}_m,\boldsymbol{\pi}_i^\star} = \left| \operatorname{acos}\left( \frac{|\boldsymbol{\eta}_m^\top \cdot \boldsymbol{\eta}_i^\star|}{\|\boldsymbol{\eta}_m\|_2 \, \|\boldsymbol{\eta}_i^\star\|_2} \right) \right| \leq 5^\circ, \qquad (1)$$

and

$$d\left(\boldsymbol{\pi}_m, \boldsymbol{\pi}_i^\star\right) = \left| d\left(\boldsymbol{\pi}_m, \mathbf{C}\right) - d\left(\boldsymbol{\pi}_i^\star, \mathbf{C}\right) \right| \leq 50 \text{ [cm]}. \quad (2)$$

The concentration of the green distribution around small distance values in Figure 2, which has been generated from $M = 200000$ planes and with 7 ground-truth planes on the well-known Herz-Jesu-P8 dataset [33], confirms the appropriateness of the proposed plane quality metric $q\left(\boldsymbol{\pi}_m\right)$.

## 3. Illustration of the matching inaccuracy and ambiguity

Figures 3(c) and 3(d) illustrate the matching inaccuracy $m_i\left(\mathbf{X}\right)$ and the matching ambiguity $m_a\left(\mathbf{X}\right)$ of the 3D points in the dense point cloud associated to the well-known Castle sequence [33].

## 4. Analysis of the scattering function

To analyze the scattering of the 3D points in $\mathcal{C}^{\boldsymbol{\tau}}_{\mathcal{R},\boldsymbol{\pi}}$ around the investigated plane $\boldsymbol{\pi}$, we have defined in Section 4 $f_{\mathcal{C}^{\boldsymbol{\tau}}_{\mathcal{R},\boldsymbol{\pi}}}\left(l, \boldsymbol{\pi}\right)$ as a function describing the fraction of 3D points in $\mathcal{C}^{\boldsymbol{\tau}}_{\mathcal{R},\boldsymbol{\pi}}$ whose distance to $\boldsymbol{\pi}$ is smaller than $l \in \mathbb{R}^+$, given a pair $\boldsymbol{\tau} = \{\tau_i, \tau_a\}$.

Figures 4 ((a) and (b)) and 5 ((a) and (b)) consider the Castle's 3D scene [33], and illustrate multiple examples of scattering functions $f_{\mathcal{C}^{\boldsymbol{\tau}}_{\mathcal{R},\boldsymbol{\pi}}}\left(l, \boldsymbol{\pi}\right)$ for different types of regions, different proposed 3D planes and different pairs of inaccuracy/ambiguity thresholds. In each of the (a) or (b) figures, the left column shows how the 2D region $\mathcal{R}$ (in red) projects on the second view via the manually defined planar model specified on the right of the arrow. The two other columns illustrate some scattering functions $f_{\mathcal{C}^{\boldsymbol{\tau}}_{\mathcal{R},\boldsymbol{\pi}}}\left(l, \boldsymbol{\pi}\right)$ for four pairs of investigated thresholds, ranging in $\tau_i = \{0.003; \infty\}$ and $\tau_a = \{0.25; \infty\}$. For each of the scattering curves, the upper ordinate value represents the number of 3D points in $\mathcal{C}^{\boldsymbol{\tau}}_{\mathcal{R},\boldsymbol{\pi}}$ while the abscissa represents the distance $l$ (in meters) to the investigated plane $\boldsymbol{\pi}$. The analysis of the scattering function is limited to $l \in [0; l_{\text{lim}}]$, where $l_{\text{lim}} = 1$[m] has been empirically chosen accordingly to the scale of the Castle's 3D scene.

Figure 4(a) considers the approximation of the roof of the Castle by the (ground-truth) 3D plane of the Castle's left wall. It is worth noticing that there is only a small percentage of the 3D points in $\mathcal{C}^{\boldsymbol{\tau}}_{\mathcal{R},\boldsymbol{\pi}}$ that are close to this plane. This can be observed in the top-middle curve in Figure 4(a), which indicates that more than 50% of the 3D points associated to the 2D roof region, are more than 1 meter away from the 3D wall plane. This observation is common to most of the cases for which the proposed plane does not correctly represent the 3D of the investigated region. Hence, a small area under the scattering curve appears to be a good indicator of the 3D plane incorrectness.

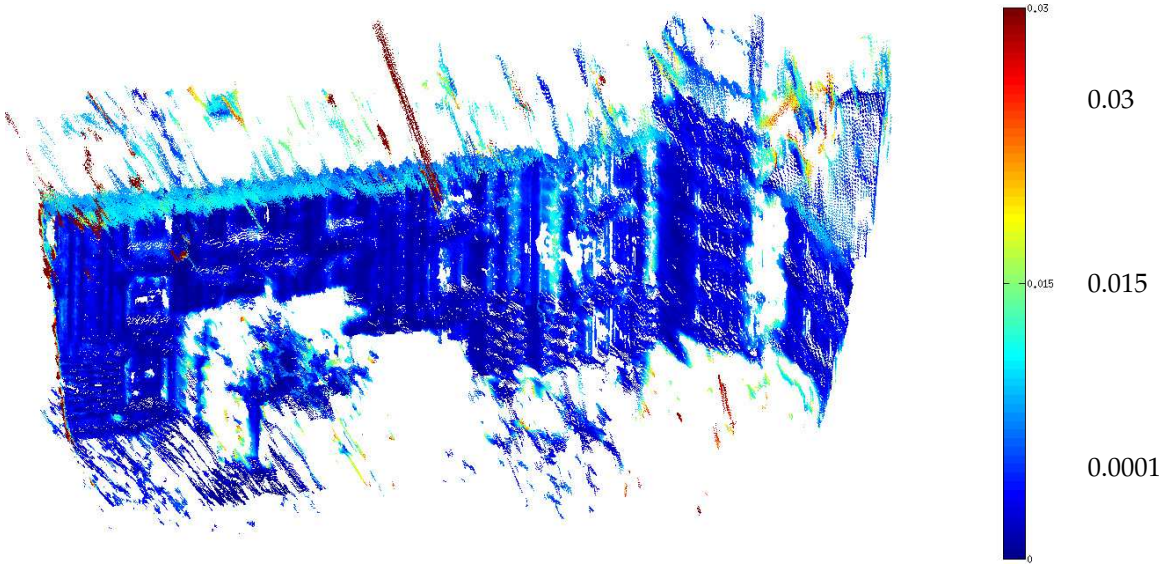
Figure 4(b) illustrates the 3D approximation of the same roof region, but this time by the (ground-truth) 3D plane associated to the roof. The areas under the scattering curves are now larger than in Figure 4(a), which confirms that this value might be a good indicator of the plane correctness. A deeper analysis of Figure 4(b) also reveals that there might be a significative amount of 3D points that are still far away from this ground-truth plane. In particular, the top-right extremity of the top-middle curve shows that only barely more than 70% of the 3D points associated to this region are closer than 1 meter from this plane. This is explained by the low precision of the 3D points representing the roof region, due to its repetitive pattern nature, which makes the matching phase error-prone. However, when restricting ourselves to the more accurate and less ambiguous 3D points, *e.g.*, $\tau_i = 0.003$ and $\tau_a = 0.25$ (Figure 4(b), curve in the bottom-right corner), we can observe that most of the 3D points are close from the ground-truth 3D plane.

Figure 5(a) investigates the modeling of a 2D region $\mathcal{R}$ of the ground-plane (floor) by the planar model of the left wall of the Castle's sequence. Although the investigated wall plane is perpendicular to the (ground-truth) floor plane, the region $\mathcal{R}_{\boldsymbol{\pi}}$ is also located on the floor. This would make any kind of conventional reprojection error reasonably small, since the texture is relatively uniform on all the floor. In contrast, the (areas under the) scattering curves depicted in Figure 5(a) reveal that the plane is not a valid model, as opposed to the one proposed in Figure 5(b).

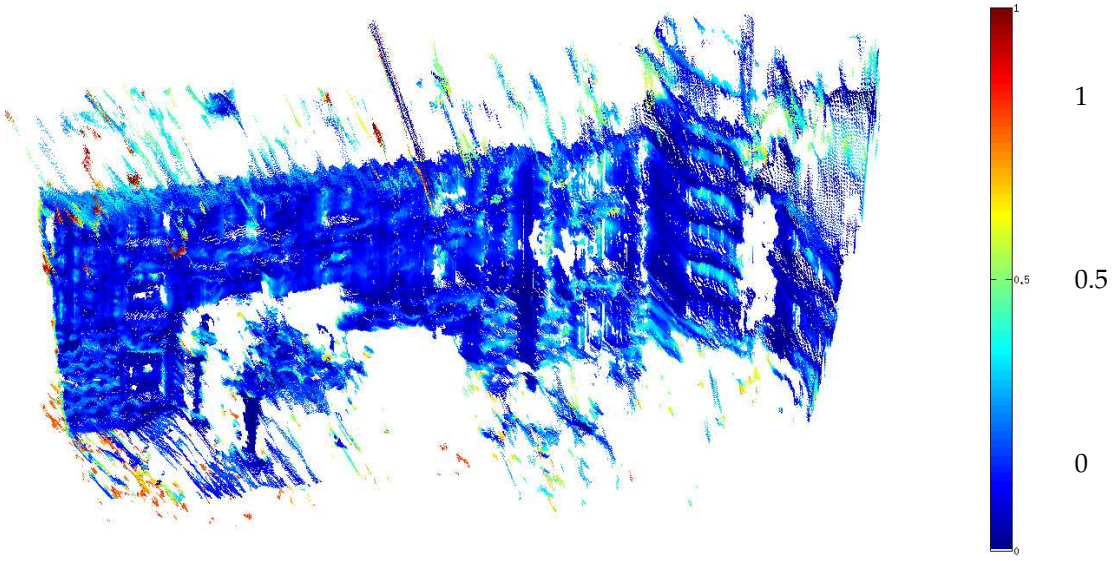
We conclude from these observations that the reliability of a proposed 3D plane can reasonably be inferred based on the analysis of the area under curve (AuC) observed for multiple sets of thresholds $\{\tau_i, \tau_a\}$.

(a) Left view

(b) Right view




(c) Matching inaccuracy $m_i(\mathbf{X})$ of the 3D points $\mathbf{X}$.




(d) Matching ambiguity $m_a(\mathbf{X})$ of the 3D points $\mathbf{X}$.

Figure 3: The uncertainty of each 3D point $\mathbf{X}$ is quantified based on (c) the matching inaccuracy $m_i(\mathbf{X})$ and on (d) the matching ambiguity $m_a(\mathbf{X})$.

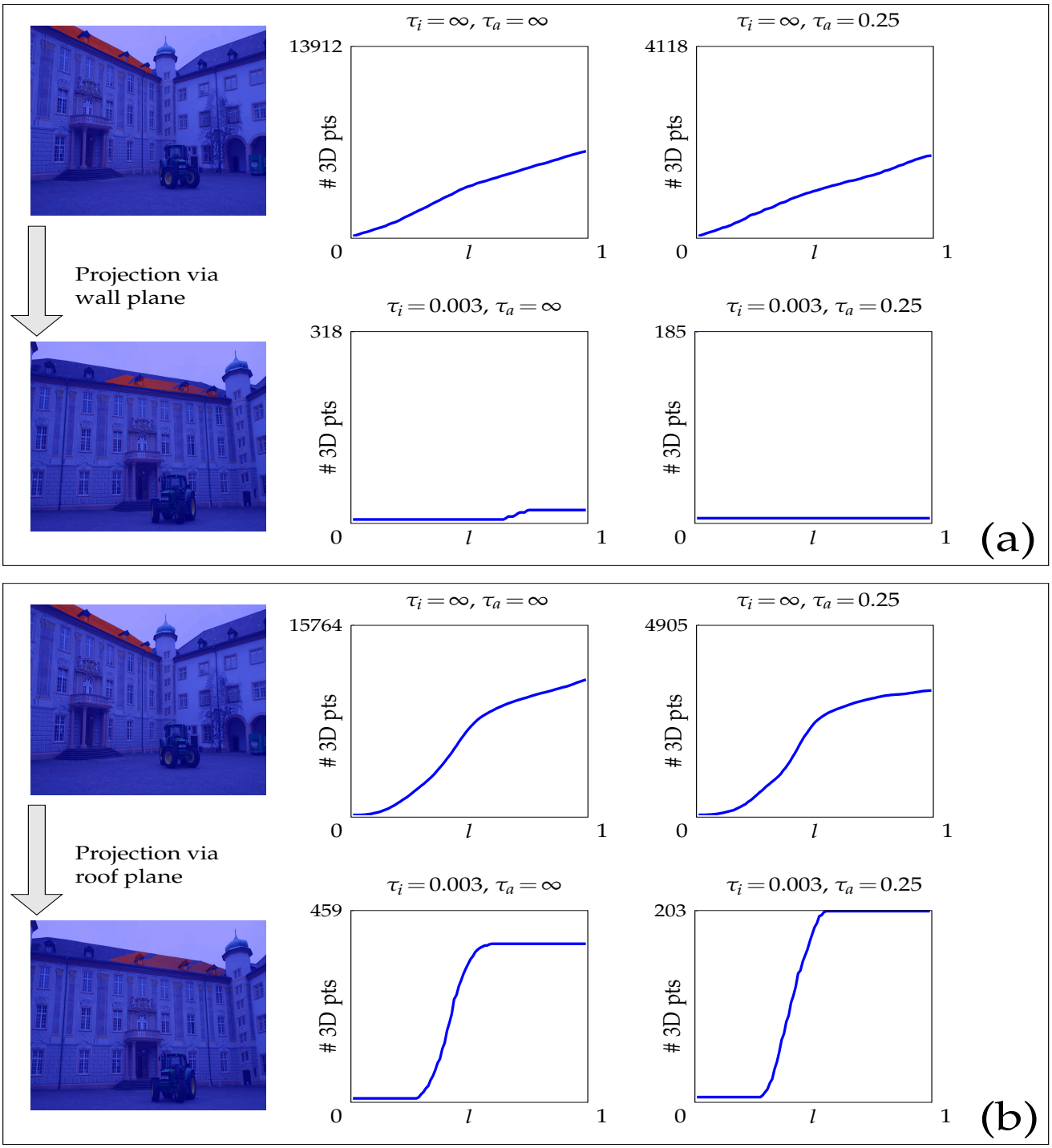



Figure 4: Our data-fidelity term considers the scattering (blue curves) of 3D points around the investigated plane $\pi$. Generally, when the investigated 3D plane does not represent correctly the region (see (a)), the proportion of 3D points in $\mathcal{C}^{\tau}_{\mathcal{R},\pi}$ near the investigated plane decreases when progressively considering more accurate and unambigious 3D points (lower values of $\tau_i$ and $\tau_a$). When the investigated 3D plane represents correctly the region (see (b)), this proportion tends to increase especially when some error is tolerated in the distance to the plane (typically when $l > 30$ [cm] in the plots).

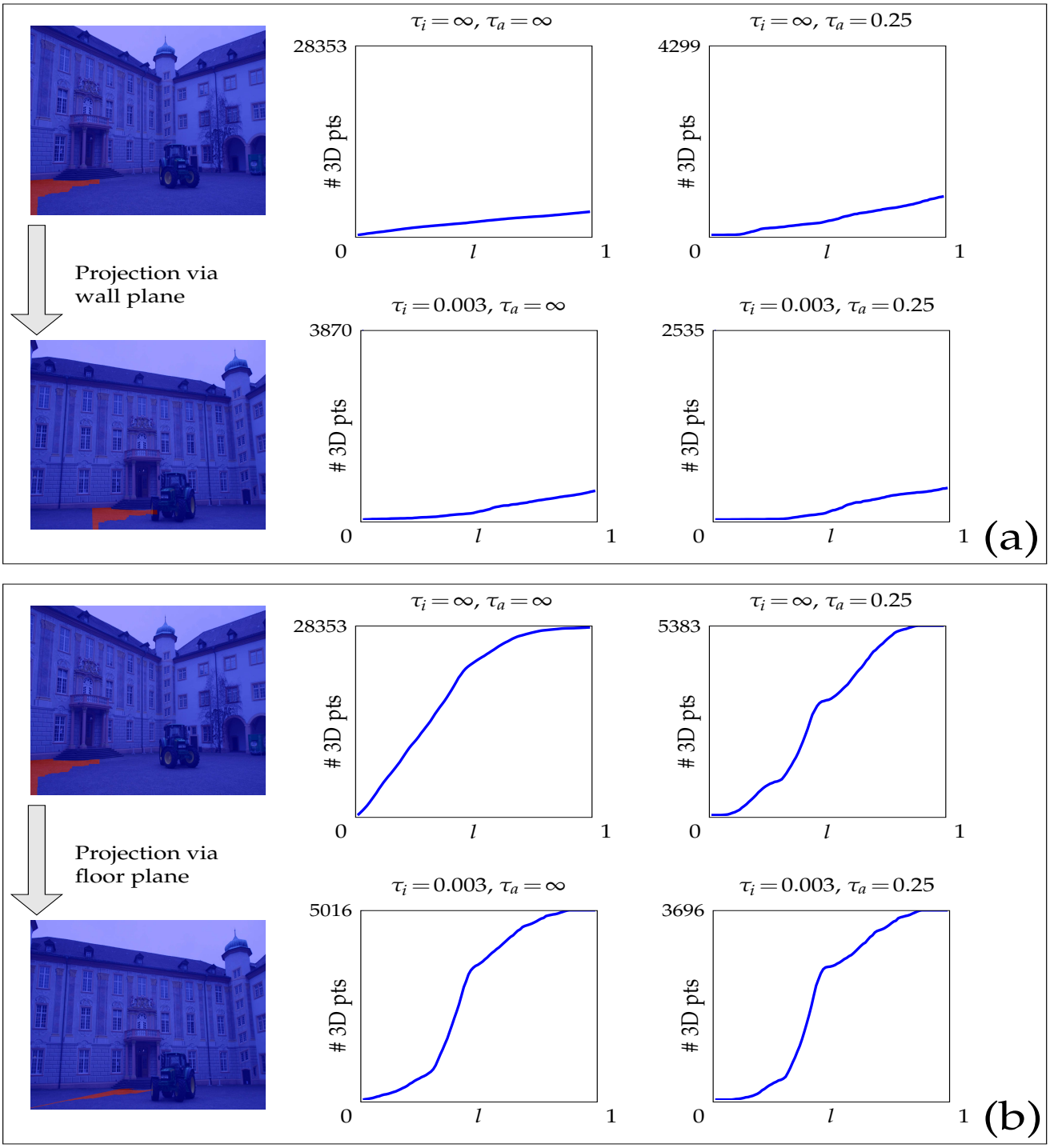
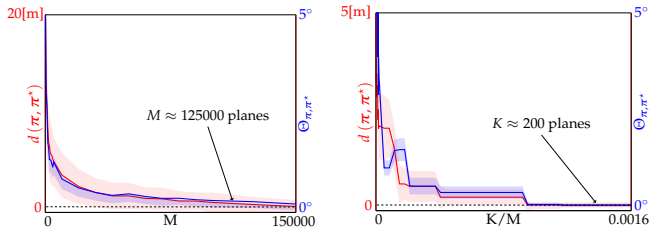



Figure 5: Evolution of the scattering function when a textureless region, *e.g.*, a floor region, is projected via the wall plane (90° away from the correct plane model). Although any data-fidelity based on an usual reprojection error would promote the plane as correct (because of the low image difference between the projected image and the image really captured), the very small areas under the scattering functions in (a) indicate that the 3D plane of the wall does not approximate correctly the 3D surface of the floor region. At the opposite, the high areas under the scattering functions in (b) indicate that the 3D plane of the floor correctly approximates the 3D surface of the floor region.

## 5. Efficiency of the plane proposition phase

To assess the relevance of our plane proposition method, Figure 6(a) first depicts, as a function of $M$, how well a set of $M$ plane candidates approximates a ground-truth 3D surface described by $I^\star$ ground-truth planes. Precisely, from each tested set of $M$ plane candidates, we have selected the $I^\star$ planes candidates approximating the best the ground-truth planes $\{\boldsymbol{\pi}_i^\star\}_{i=\{1,\cdots,I^\star\}}$, *i.e.*, the ones minimizing $\|\boldsymbol{\eta}_i - \boldsymbol{\eta}_i^\star\|_2$ with $\boldsymbol{\eta}_i$ and $\boldsymbol{\eta}_i^\star$ being defined as in Section 3.




(a) (b)

Figure 6: (a) Average of the maximum remoteness, with respect to a set of ground-truth planes $\{\boldsymbol{\pi}_i^\star\}_{i=\{1,\cdots,I^\star\}}$, of the best plane candidates $\boldsymbol{\pi}_m$ among the $M$ plane candidates. (b) Average of the maximum remoteness of the best proposed plane $\boldsymbol{\pi}_k$ among the proposed $K$ planes with respect to a set of ground-truth planes $\{\boldsymbol{\pi}_i^\star\}_{i=\{1,\cdots,I^\star\}}$.

The blue and red curves in Figure 6(a) respectively plot the worst relative angular and distance errors (cfr. Equations (1) and (2) in this supplementary material), as measured with respect to $I^\star = 7$ ground-truth planes of the well-known Herz-Jesu-P8 dataset [33]. The performances have been averaged on $P = 1000$ iterations of the random triplets-based selection of the $M$ planes (cfr. Section 3 in the paper). The blue and red shaded areas in this figure illustrate respectively the standard deviations of these two metrics.

From Figure 6(a) , we observe that multiple plane candidates ($M \geq 125000$) are required to ensure that the set of $M$ candidate planes represent "correctly"[7] the ground-truth 3D surface.

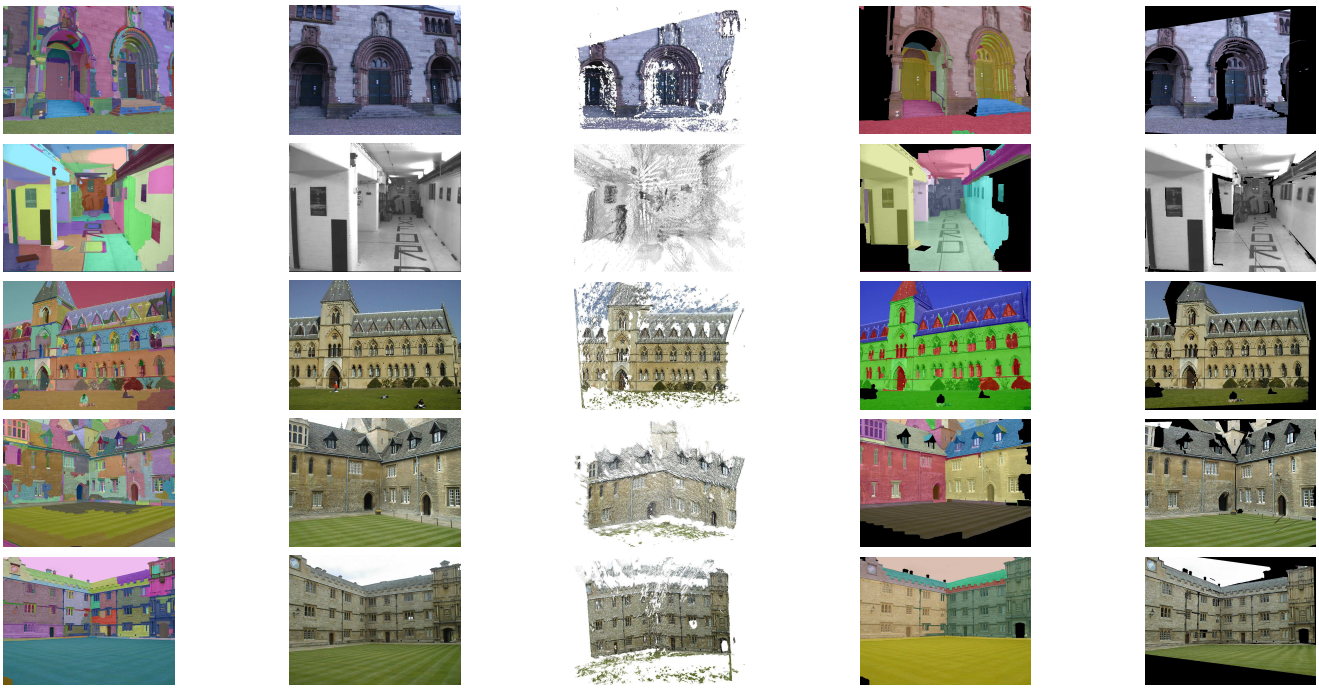
Figure 6(b) depicts the same angular and distance errors as a function of $K$, when $M = 125000$. From this figure, we observe that our plane proposition phase effectively proposes a highly limited set of 3D plane ($K$ is set to 200 in all our experiments), which greatly reduces the number of hypotheses (in practice, $M$ had to be chosen 2 to 3 orders of magnitude higher than $K$ to make sure to include all the ground-truth plane models of the scene) by eliminating redundant ones. This allows the PEARL optimization to work efficiently on a reduced set of proposed planes.

[7] Even if the plane candidates do not perfectly correspond to the ground planes, their parameters are later optimized in the "re-learn" phase of PEARL (see Section 5 in the paper).

## 6. 3D reconstruction and view interpolation experiments

Figure 7 completes Figure 2 (in the paper) by presenting the results of the different steps of our algorithm on the other well-known wide-baseline datasets, namely (from top to down) the HerzjesuP25 [33], Oxford Corridor/Library/Merton I and II [41] datasets. Despite the presence of noise in the dense 3D point cloud (column (c)), the 3D planes are correctly estimated. We attribute the high performances of our algorithm to the robustness of the newly proposed data-fidelity term, which incorporates the matching accuracy and the matching ambiguity into a new 3D fitting error, instead of simply combining a conventional fitting error with a region-based projection error (*e.g.*, based on a weighted average whose ad-hoc weights are application dependent).

Figure 8 completes the virtual views proposed in Figure 3, by projecting the (textured) 3D piecewise-planar model obtained for the CastleP19/FountainP11 [33] and Model-house/Wadham/MertonIII [41] datasets. Videos representing these transitions are attached to this supplementary material.



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

Figure 7: (Best viewed in color). Based on the segmentation (a) of one of the two wide-baseline views ((a) and (b)) and on their associated dense point cloud (c), our method approximates the 3D surface by the minimum set of of 3D planes. In (d), regions assigned to the same 3D plane are illustrated with a same color. The reprojection of the optimal piecewise-planar reconstruction, textured based on the first view (a) and projected in the second view (b), is represented in (e).

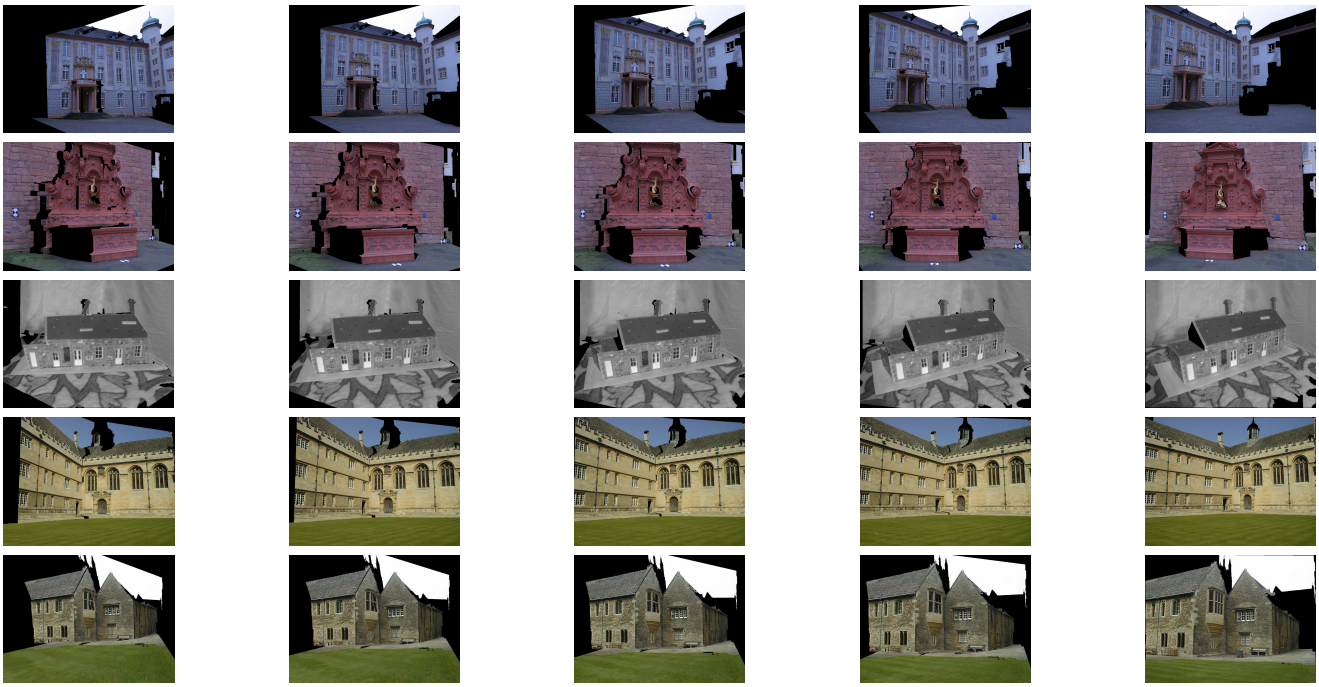


Figure 8: Projection of the textured piecewise-planar approximation of the scene's 3D on virtual views in-between the two cameras of the wide-baseline stereo pair.