Image Quality Assessment for Photo-consistency Evaluation on Planar Classification in Urban Scenes

Marie-Anne Bauda, Sylvie Chambon, Pierre Gurdjos, Vincent Charvillat

Image Quality Assessment for Photo-consistency Evaluation on Planar Classification in Urban Scenes

M.-A. Bauda[1,2], S. Chambon[1], P. Gurdjos[1] and V. Charvillat[1]
[1]*VORTEX, University of Toulouse, IRIT-ENSEEIHT, Toulouse, France*
[2]*imajing sas, Ramonville St Agne, France*
{*mbauda, schambon, pgurdjos, charvi* }*@enseeiht.fr*

Keywords: Image segmentation, Urban Scene, Planar classification, Image Quality Assessment.

Abstract: In the context of semantic segmentation of urban scenes, the calibrated multi-views and the flatness assumption are commonly used to estimate a warped image based on the homography estimation. In order to classify planar and non-planar areas, we propose an evaluation protocol that compares several Image Quality Assessments (IQA) between a reference zone and its warped zone. We show that cosine angle distance-based measures are more efficient than euclidean distance-based for the planar/non-planar classification and that the Universal Quality Image (UQI) measure outperforms the other evaluated measures.

1 INTRODUCTION

The semantic segmentation consists in detecting and identifying objects present in the scene. For example, in urban scenes, we would like to distinguish the ground (road, pavement) from the façades of the buildings. A first step for solving this problem consists in using an over-segmentation, such as superpixel construction. It is an intermediate feature of interest, in comparison with using pixels only or with the use of regular patches, that combines a space support and a photometric criterion (Felzenszwalb and Huttenlocher, 2004; Achanta et al., 2012). These methods aim at facilitating the segmentation and they allow to pre-process high resolution images by reducing the problem complexity (Arbelaez et al., 2009).

Illustrated in figure 1, the two superpixels SP1 and SP2 are helpful for the semantic segmentation because they are coherent with the scene geometry. However, the striped non-planar superpixel SP3, is not well adapted because it is astride a boundary of two adjacent planes, i.e. two façades. A superpixel should represent a meaningful 3D surface.

Regarding urban scenes, planar geometry constraints are commonly used as prior knowledge on the context in monocular images (Saxena et al., 2008; Hoiem et al., 2008; Gould et al., 2008). An intermediate level of image segmentation is to classify zones into planar and non-planar classes but the choice of a discriminative similarity measure remains difficult. If multiple images are available, the sparse

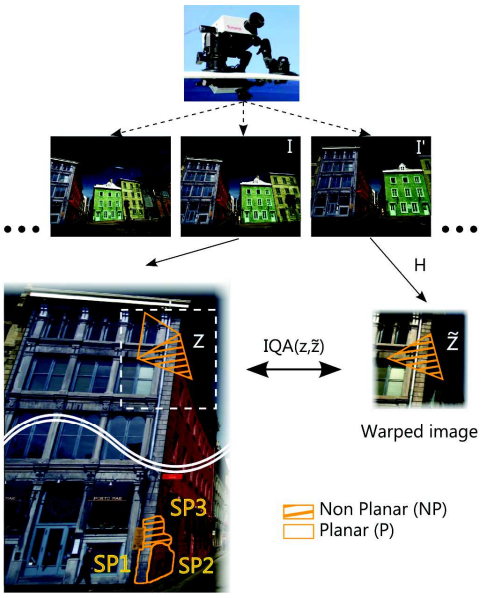
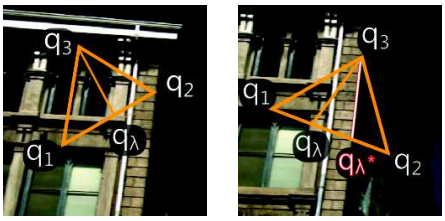


Figure 1: Superpixel analysis and presentation of the IQA evaluation protocol – It is based on a photo-consistency criterion IQA between a piece of the reference image $\mathbf{z}$ and its corresponding warped area $\tilde{\mathbf{z}}$ estimated by the homography H induced by the plane of support.

3D point clouds and the epipolar geometry are useful to strengthen the understanding of the scene (Bartoli, 2007; Mičušík and Košecká, 2010; Gallup et al., 2010).

In particular, under the planar hypothesis, know-

ing the epipolar geometry and the orientation of the represented surface, the homography estimation between two regions is defined. Then, an Image Quality Assessment (IQA) is used to evaluate the similarity (or the dissimilarity) between the initial area $\mathbf{z}$ and the warped area $\tilde{\mathbf{z}}$ from which we can deduce the planarity of $\mathbf{z}$. The IQA$(SP1, \tilde{SP1})$ and the IQA$(SP2, \tilde{SP2})$ are more similar than IQA$(SP3, \tilde{SP3})$, cf. Figure 2 that shows an example of this behaviour with a planar and a non-planar regions $\mathbf{z}$ delimited by three 2D points noted q_1, q_2 and q_3.



(a) P (b) NP

Figure 2: Two regions of interest $\mathbf{z}$ of the reference image I: (a) one planar and (b) one non-planar. The point q_λ follows the line $[q_1 q_2]$. The intersection of the two planes $\pi_1 \cap \pi_2$ is denoted $q_{\lambda^\star}$ which corresponds to our ground truth, i.e. it delimits the edge between the two planes.

In this work, two successive calibrated images I and I' are used, cf. Figure 3. We denote $\mathtt{P}_I = \mathtt{K}[\mathtt{I}|\mathbf{0}]$ the projection matrix of I, where $\mathtt{K}$ is the matrix of the intrinsic parameters and $\mathtt{P}_{I'} = \mathtt{K}[\mathtt{R}|\mathbf{t}]$ the projection matrix associated to the image I' where $\mathtt{R}$ is the rotation matrix and $\mathbf{t}$ the translation vector that determines the relative poses of the cameras. Each 3D point Q_i corresponds to 2D matched points $q_i \in I$ and $q'_i \in I'$.

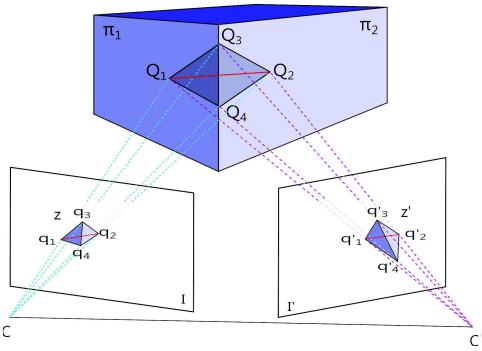


Figure 3: Configuration between I and I': the 3D points Q_i obtained by the 2D matched points $q_i \leftrightarrow q'_i$. It determines the regions of interest $\mathbf{z}$ and $\mathbf{z}'$ to estimate the homography.

In order to obtain an over-segmentation consistent with the scene geometry, in this article, we use the flatness assumption on objects represented in the images to compute the similarity between $\mathbf{z}$ and $\tilde{\mathbf{z}}$. One of the difficulty of this problem is to choose a pertinent measure between several similarity and dissimilarity measures used in the literature such as the r-consistency in (Kutulakos, 2000; Bartoli, 2007) or the Zero mean Cross Correlation (ZNCC) (Quan et al., 2007; Häne et al., 2013). Therefore, we propose an IQA evaluation protocol for planar/non-planar regions classification. It means that we want to highlight the measure that is the most sensitive to photometric differences induced by the estimation of the warped corresponding areas in non-planar case. To simplify the problem and as a preliminary work, we apply IQA over triangles instead of superpixels because only three matched points are required to estimate the homography if the epipolar geometry of the two views is known.

The next section presents five of the existing measures. We also introduce a new measure that merges two main ideas of two existing measures. Then, in § 3, our evaluation protocol is detailed. Finally, experimentations followed by our analyses are presented.

2 PHOTO-CONSISTENCY MEASURES

The quantification of how a reference region $\mathbf{z}$ and a target region $\tilde{\mathbf{z}}$ are photometrically similar (or dissimilar) is computed from photo-consistency measures which compare pixel intensities. Here, source and target regions are delimited by the same polygon, the former includes pixels of the reference image while the latter includes warped pixels obtained by transferring intensities from the target image, under the planarity assumption of the projected surface.

We propose a classification of photo-consistency measures into two classes: euclidean distance-based or cosine angle distance-based measures. All the measures are illustrated in figure 4 and we note:

- $N = \text{card}\{q_i \in \mathbf{z}\}$ is the number of pixels in the considered region $\mathbf{z}$ (or equivalently $\tilde{\mathbf{z}}$);
- v_i (resp. $\tilde{v}_i$) is the luminance coordinate in CIELab color space of pixel q_i in the region $\mathbf{z}$ (resp. $\tilde{\mathbf{z}}$).

Euclidean Distance-based Measures – Denoted by IQA$_d$, they quantify the *dissimilarity* between the region $\mathbf{z}$ and the warped corresponding region $\tilde{\mathbf{z}}$ by relying on the Euclidean distance between the two vectors v and $\tilde{v}$ linearising $\mathbf{z}$ and $\tilde{\mathbf{z}}$. The first one is the well-known Mean Square Error (MSE), defined by:

$$\text{MSE}(\mathbf{z}, \tilde{\mathbf{z}}) = \frac{1}{N} \sum_i (v_i - \tilde{v}_i)^2 \qquad (1)$$

This measure can be extended, if for a given pixel q_i the square neighbourhood in a radius less than r is

considered:

$$\mathrm{MSE}_r(\mathbf{z},\tilde{\mathbf{z}}) = \frac{1}{N}\sum_i [\frac{1}{(2r)^2} \sum_{j\,/\,|q_i - q_j| \leq r} (v_j - \tilde{v}_j)^2] \quad (2)$$

The r-consistence used in (Bartoli, 2007) also falls into this category. For a given pixel $q_i \in \mathbf{z}$, the pixel difference in the r-ring neighbourhood of the correspondent pixel $q'_i \in \mathbf{z}'$ is searching.

$$\mathrm{RC}_r(\mathbf{z},\tilde{\mathbf{z}}) = \frac{1}{N}\sum_i \Big(\min_{j\,/\,(q_i - q_j)^2 < r^2} |v_i - \tilde{v}_j| \Big)^2 \quad (3)$$

Cosine Angle Distance-based Measures – Denoted by IQA_s, they quantify the *similarity* of the regions by relying on the inner product of the two vectors. Typically, these vectors are treated as random variables and "a correlation coefficient" is computed by dividing the covariance of the two variables by the product of their standard deviations.

In this work, we will consider the Structural SIMilarity (SSIM) coefficient (Wang et al., 2004). Three statistical terms are involved: luminosity $l(\mathbf{z},\tilde{\mathbf{z}})$, contrast $c(\mathbf{z},\tilde{\mathbf{z}})$ and structure $s(\mathbf{z},\tilde{\mathbf{z}})$. Moreover, Gaussian weights are introduced to give more importance to the central pixel. If the following terms are defined:

- $\mu_{\mathbf{z}}$ (resp. $\mu_{\tilde{\mathbf{z}}}$) the mean of v_i (resp. $\tilde{v}_i$) over the region $\mathbf{z}$ (resp. $\tilde{\mathbf{z}}$),
- $\sigma_{\mathbf{z}}$ (resp. $\sigma_{\tilde{\mathbf{z}}}$) the standard deviation of $\mathbf{z}$ (resp. $\tilde{\mathbf{z}}$),
- $\sigma_{\mathbf{z}\tilde{\mathbf{z}}}$ the covariance of $\mathbf{z}$ and $\tilde{\mathbf{z}}$,

then, the SSIM is defined by:

$$\mathrm{SSIM}(\mathbf{z},\tilde{\mathbf{z}}) = l(\mathbf{z},\tilde{\mathbf{z}}) \cdot c(\mathbf{z},\tilde{\mathbf{z}}) \cdot s(\mathbf{z},\tilde{\mathbf{z}}) \quad (4)$$

where:

$$l(\mathbf{z},\tilde{\mathbf{z}}) = \frac{2\mu_{\mathbf{z}}\mu_{\tilde{\mathbf{z}}} + \alpha}{\mu_{\mathbf{z}}^2 + \mu_{\tilde{\mathbf{z}}}^2 + \alpha}, \qquad c(\mathbf{z},\tilde{\mathbf{z}}) = \frac{2\sigma_{\mathbf{z}}\sigma_{\tilde{\mathbf{z}}} + \beta}{\sigma_{\mathbf{z}}^2 + \sigma_{\tilde{\mathbf{z}}}^2 + \beta}$$

$$\text{and} \quad s(\mathbf{z},\tilde{\mathbf{z}}) = \frac{\sigma_{\mathbf{z}\tilde{\mathbf{z}}} + \gamma}{\sigma_{\mathbf{z}}\sigma_{\tilde{\mathbf{z}}} + \gamma}.$$

The constants α, β and γ are introduced to avoid to divide by zero. This case occurs when a region is homogeneous in intensity, in that case $\sigma_{\mathbf{z}} = 0$ or when it is a black zone, i.e. $\mu_{\mathbf{z}} = 0$. SSIM is symmetric, stacked and reaches its maximum when the two areas are similar i.e. $\mathbf{z} = \tilde{\mathbf{z}}$. Let us remark that the structure term $s(\mathbf{z},\tilde{\mathbf{z}})$ corresponds to the Zero mean Normalised Cross-Correlation (ZNCC) for $\gamma = 0$ (Aschwanden and Guggenbül, 1992). The Universal Quality Index (UQI) (Z. Wang and Bovik, 2002) corresponds to the special case where $\alpha = \beta = \gamma = 0$ and without any weight balancing. This means that all pixels in the sliding window have the same importance. More precisely, UQI is formulated as follow:

$$\mathrm{UQI}(\mathbf{z},\tilde{\mathbf{z}}) = \frac{4\sigma_{\mathbf{z}\tilde{\mathbf{z}}}\,\mu_{\mathbf{z}}\mu_{\tilde{\mathbf{z}}}}{(\sigma_{\mathbf{z}}^2 + \sigma_{\tilde{\mathbf{z}}}^2)\,[\mu_{\mathbf{z}}^2 + \mu_{\tilde{\mathbf{z}}}^2]} \quad (5)$$

We proposed here a new metric called RUQI, combining ideas from UQI by using statistic over the r-neighbourhood and by optimising the similarity on a neighbourhood such as in RC_r:

$$\mathrm{RUQI}(\mathbf{z},\tilde{\mathbf{z}}) = \frac{1}{N}\sum_i \Big(\max_{j\,/\,(q_i - q_j)^2 < r^2} (\mathrm{UQI}(\xi_i, \tilde{\xi_j})) \Big) \quad (6)$$

where ξ_i (resp. $\tilde{\xi_j}$) is defined as a small window of $\mathbf{z}$ (resp. $\tilde{\mathbf{z}}$) around q_i (resp. q_j).

Conclusions and analyses about IQA – First of all, regarding the IQA when $r = 0$ then $\mathrm{RC}_0 = \mathrm{MSE}_0 = \mathrm{MSE}$. In details, RC_r optimizes the difference on the warped image and MSE compares pixel-to-pixel while MSE_r takes into account the neighbourhood.

In SSIM, a Gaussian weight is used to give more importance to the central pixel, compared to UQI measure. So, we introduce RUQI that optimizes the similarity in the r-neighbourhood over the statistical information to combine advantages of UQI and RC.

Finally, the cosine angle distance-based measures compute statistics over pixels belonging to the zone instead of a simple difference. All these measures introduce a parameter r and we discuss the influence of this parameter in § 4.

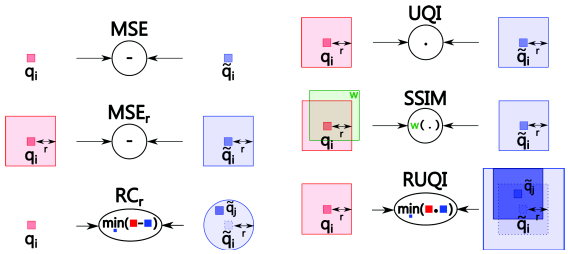


Figure 4: $\mathrm{IQA}(\mathbf{z},\tilde{\mathbf{z}})$ computation (Euclidean distance-based/cosine angle distance-based measures) on a reference zone $\mathbf{z}$ centred on q_i (in red) and on $\tilde{\mathbf{z}}$ centred on $\tilde{q}_i$ (in blue). In RC_r and RUQI, the point $\tilde{q}_j$ corresponds to the more similar pixel in a r-neighbourhood, see § 2 for details.

3 IQA PROTOCOL EVALUATION

Assuming that a high photo-consistency is obtained when correct surface orientations are known, this IQA evaluation protocol highlights the measure that fits with the assumption to discriminate planar from non-planar regions. More precisely, we want

to answer the following question: if a triangle is supported by two planes, can we, by estimating homographies, detect the plane switching with a photometric criterion? We would like to detect non-planar regions in order to cut it until obtaining planar regions. To make easier the evaluation task, it is natural to introduce this simplification: one of the vertex lies on the intersection.

For this reason, our approach is based on the main idea that when $\lambda \in [0-1]$, we are expecting to a constant and high similarity (low dissimilarity) curve in a planar case, and in a non-planar case for low similarity (high dissimilarity). When a region of interest is supported by two planes, an extrema can be reach at the intersection of both planes, the ground truth, noted $\lambda^\star$. An overview of our proposed approach is presented in algorithm 1.

3.1 Homographies Estimation

To compare zones, we estimate homographies induced by the plane supports, to compute the warped image $\tilde{\mathbf{z}}$ from $\mathbf{z}'$. First, we split the region $\mathbf{z}$ defined by $q_1q_2q_3$ into two smaller triangles $q_1q_3q_\lambda$ and $q_2q_3q_\lambda$. Shown in figure 2, the point q_λ lies on the segment $[q_1q_2]$ and is defined by $q_\lambda = \lambda q_1 + (1-\lambda)q_2$ where $\lambda \in [0,1]$ in I. Since we have a perspective transformation between the two views, $q'_\lambda \neq \lambda q'_1 + (1-\lambda)q'_2$. Therefore, four 2D matched points or three 2D matched points and epipoles allow to estimate the homography induced by the 3D plane. To compute this transformation, a few methods provided by (Hartley and Zisserman, 2004) exist and it is achieved by the $computeHomography(.)$ function over three matched points.

The 3D points Q_1 and Q_2 stand on each plane π_1 and π_2, as shown in figure 3. The points Q_3 and Q_4 are located on the edge between the two planes. The interested area in the referenced (resp. adjacent) image $\mathbf{z}$ (resp. $\mathbf{z}'$), is defined by the triangle $q_1q_2q_3$ (resp. $q'_1q'_2q'_3$). The homography H_1 is induced by the plane support π_1 and is well defined if none of the three points are aligned. The H_1 enables us to estimate $\tilde{z}_1$ defined by the projection of the adjacent region of interest on the reference image. The same goes for H_2 and the plane π_2. In consequence, we have:

$$\tilde{\mathbf{z}}_k = \{\tilde{q}_i = H_k q'_i \;/\; q'_i \in \mathbf{z}'_k\} \quad \text{where} \quad k \in \{1,2\}.$$

With a correct positioning of q'_λ we can adjust the homography estimation. Once homographies are estimated, the warped zone $\tilde{\mathbf{z}}$ is obtained by interpolating the zone $\mathbf{z}'$ through the homography transformation.

Data: 4 matching points of interest $q_1 \leftrightarrow q'_1$, $q_2 \leftrightarrow q'_2$, $q_3 \leftrightarrow q'_3$, $q_4 \leftrightarrow q'_4$ over two images I and I'
Result: Planar/non-planar classification

// Estimation of right value $\lambda^\star$
$q_{\lambda^\star} \leftarrow (q_1q_2) \cap (q_3q_4)$;
// Estimation of homographies (§ 3.1)
$H_1 \leftarrow computeHomography(q_3, q_4, q_1)$;
$H_2 \leftarrow computeHomography(q_3, q_4, q_2)$;
// Computation of IQA value for each λ
for $\lambda = 0 : d\lambda : 1$ **do**
  // Computation of the point $\mathbf{q}_\lambda \in [\mathbf{q}_1\mathbf{q}_2]$
  $q_\lambda \leftarrow \lambda q_1 + (1-\lambda)q_2$;
  // Estimation of the warped image
  if $\lambda < \lambda^\star$ **then**
    $q_{\lambda'} \leftarrow H_1(q_\lambda)$;
    $H \leftarrow computeHomography(q_2, q_3, q_\lambda)$;
    $\tilde{\mathbf{z}}_1 \leftarrow H_1(\mathbf{z}')$;
    $\tilde{\mathbf{z}}_2 \leftarrow H(\mathbf{z}')$;
  else
    $q_{\lambda'} \leftarrow H_2(q_\lambda)$;
    $H \leftarrow computeHomography(q_1, q_3, q_\lambda)$;
    $\tilde{\mathbf{z}}_1 \leftarrow H(\mathbf{z}')$;
    $\tilde{\mathbf{z}}_2 \leftarrow H_2(\mathbf{z}')$;
  end
  $\tilde{\mathbf{z}} \leftarrow \tilde{\mathbf{z}}_1 \cup \tilde{\mathbf{z}}_2$;
  // Computation of the IQA value (§ 3.2)
  $IQA(\lambda, \mathbf{z}, \tilde{\mathbf{z}}) \leftarrow computeIQA(\mathbf{z}, \tilde{\mathbf{z}})$;
  // Classification in P/NP region (§ 3.3)
  if $max\big(\, IQA_s(z, \tilde{z}) \,\big) > \varepsilon$ **then**
    $\mathcal{C}(\mathbf{z}, \tilde{\mathbf{z}}) \leftarrow$ P;
  else
    $\mathcal{C}(\mathbf{z}, \tilde{\mathbf{z}}) \leftarrow$ NP;
  end
end

Algorithm 1: Proposed IQA evaluation protocol applies on P/NP classification. All the steps are developed in the section 3.

3.2 IQA Computation

The comparison between $\mathbf{z}$ and $\tilde{\mathbf{z}}$ is done with IQA presented § 2. In algorithm 1, the IQA values are computed by $computeIQA(\mathbf{z}, \tilde{\mathbf{z}})$ for each pixel of the zone. They can be integrated over each pixel to merge information for each λ. Example of obtained results on a non-planar region is shown at figure 5.

3.3 Planar Classification (P/NP)

To evaluate the influence of the IQA on the quality of the classification, we use a simple classification ap-

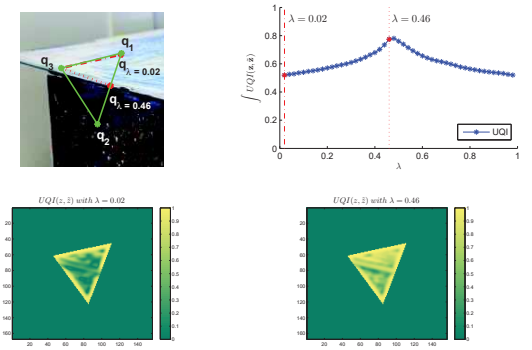


Figure 5: Example of UQI on a NP zone. At the top: the region $\mathbf{z}$ from the reference image I and the UQI means curve depending on λ. At the bottom: $\mathrm{UQI}(\mathbf{z},\tilde{\mathbf{z}})$ obtained over a NP zone for $\lambda = 0.02$ and for the ground truth $\lambda^{\star} = 0.46$ where surface orientations are correctly estimated.

Figure 6: Zones classified by increasing UQI values. The row 1 and 2 correspond to NP cases, and the last row is P cases. On each image pair, the maximum UQI value obtained is written at the bottom left of the zone $\mathbf{z}$.

proach: thresholding. As we want to highlight the best IQA candidate, i.e. the IQA that gives the best separation between both classes, we manually select the best threshold ε that maximizes the true positive rate. In our application, errors on planar zone have less impact on the results than errors on non-planar zone classification because it is preferable to cut a planar zone than not to cut a non-planar zone. In consequence, the classification is done by using the following expression:

$$C(\mathbf{z},\tilde{\mathbf{z}}) = \begin{cases} \text{NP} & \text{if } min(\text{IQA}_s(\mathbf{z},\tilde{\mathbf{z}})) < \varepsilon \\ \text{P} & \text{otherwise.} \end{cases} \quad (7)$$

4 EXPERIMENTATION

In our experiment we apply our IQA evaluation protocol to compare the six presented measures: MSE, MSE_5, RC_5, SSIM, UQI, RUQI presented in § 2. This is done on images from two datasets: images acquired in a control environment lighting and real outdoor urban scene images.

Database – BD1 corresponds to a box where sides are textured separately. Images are acquired in a controlled light environment. BD2 images are outdoor scene data and come from Oxford[1] public and available database and from calibrated images acquired with the mobile mapping system imajbox® from imajing[2] company, shown figure 1. 87 zones were evaluated (29 from BD1, 58 from BD2). Image resolutions are between 1224x1025 and 1024x768. The 2D points of interest are detected in each image, then they are matched to estimate a 3D position of each point which is bundle adjusted to reduce the positioning error. This kind of input data (2D and 3D points position, P_I and $P_{I'}$) can be generated from a structure from motion system which takes into account multiple images, such as VisualSfM (Wu, 2013).

Results and Analyses – All previous measures are evaluated with Precision-Recall (PR) and Receiver Operator Characteristic (ROC) curves, shown figure 7. Details and explanations of the relation between these two curves are given in (Davis and Goadrich, 2006). We have worked on data with the highest resolution available since we have remarked that lower the resolution is, more zones are similar and less discriminative IQA are.

The parameter r which corresponds to the size of the neighbourhood taken into account in MSE_r, UQI, SSIM and RUQI, influences the results in the following way: the larger r is, the less significant the IQA is. It means that there is a higher IQA value between two corresponding pixels than two mismatched pixels obtained in non-planar cases. The parameter r for RC, corresponds to the searching window for finding the best match. The higher it is, the higher the errors can be introduced. Moreover, it means that even when the zone is not planar, we will find a correspondent that gives a low IQA value. So, it will have the same behaviour as MSE when r is increased.

Cosine angle-distances use statistics over neighbourhood pixels and overcome results obtained with distance-based measures (red curves are above blue curves).

The planar classification is done in order to cut non-planar zone and to build a triangular mesh coherent with the geometry. So, the non-planar class

[1] www.robots.ox.ac.uk/~vgg

[2] www.imajing.eu

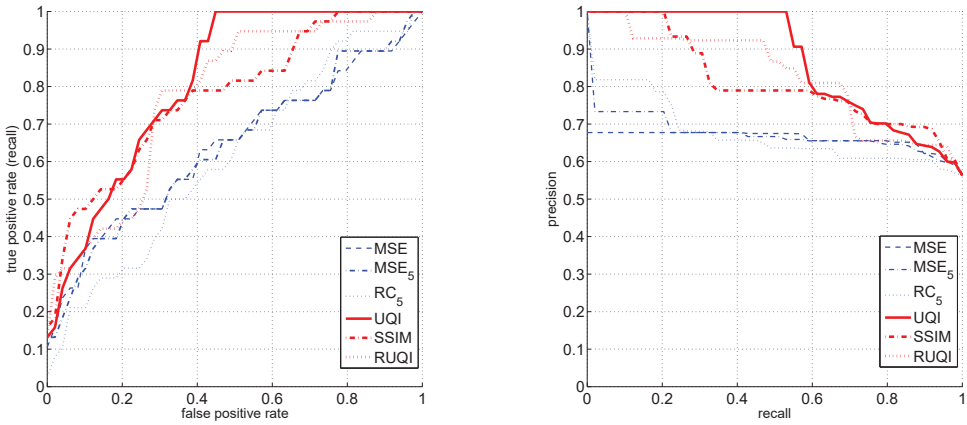


Figure 7: Obtained results on all data (6 measures on 87 triangles). On the left: the ROC and on the right: the PR curves. Dot product-based measures (red curves) are more efficient than distance-based measures (blue curves) and UQI overcomes all the others measures.

corresponds to the positive case. Best results on both classes, are obtained with UQI.

5 CONCLUSION

In order to obtain a planar/non-planar classification of zones, we have proposed an evaluation protocol which able to compare state-of-the-art of photo-consistency measures. We define a new photo-consistency measure, RUQI which combines the advantage of both UQI and RC methods.

We conclude that cosine angle distance-based are more adapted than difference-based measures for planar/non-planar classification. Among this measures, UQI overcomes other measures. Blurred images and low resolution are two limitations of our protocol, since they both induce erroneous data in the image comparison.

Our next work will consist of applying this measure in superpixel constructor to obtain a semantic segmentation taking into account the geometry of the scene through homography estimation.

REFERENCES

Achanta, R., Shaji, A., Smith, K., Lucchi, A., Fua, P., and Susstrunk, S. (2012). SLIC superpixels. In *IEEE PAMI*.

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

Aschwanden, P. and Guggenbül, W. (1992). Experimental results from a comparative study on correlation type registration algorithms. In *Robust computer vision: Quality of Vision Algorithms*.

Bartoli, A. (2007). A random sampling strategy for piecewise planar scene segmentation. In *CVIU*.

Davis, J. and Goadrich, M. (2006). The relationship between precision-recall and roc curves. In *ICML*.

Felzenszwalb, P. and Huttenlocher, D. (2004). Efficient graph-based image segmentation. In *IJCV*.

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

Gould, S., Rodgers, J., Cohen, D., Elidan, G., and Koller, D. (2008). Multi-class segmentation with relative location prior. In *IJCV*.

Häne, C., Zach, C., Cohen, A., Angst, R., and Pollefeys, M. (2013). Joint 3d scene reconstruction and class segmentation. In *IEEE CVPR*.

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

Hoiem, D., Efros, A., and Herbert, M. (2008). Closing the loop on scene interpretation. In *IEEE CVPR*.

Kutulakos, K. (2000). Approximate n-view stereo. In *ECCV*.

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

Quan, L., Wang, J., Tan, P., and Yuan, L. (2007). Image-based modeling by joint segmentation. In *IJCV*.

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

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

Wu, C. (2013). Towards linear-time incremental structure from motion. In *IEEE International Conference on 3DTV-Conference 2013*.

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