

Matching Cylindrical Panorama Sequences using Planar Reprojections

Jean-Lou De Carufel and Robert Laganière
VIVA Research lab
University of Ottawa
Ottawa, ON, Canada, K1N 6N5
jdecaruf, laganier@uottawa.ca

Abstract

This paper presents a matching scheme for large set of omnidirectional images sequentially captured in an urban environment. Most classical image matching methods when applied to cylindrical panoramas taken in large environments does not always produce a sufficient number of matches. In this work, our objective is to making sure that the full set of panoramas remains as connected as possible at all geographical locations even if only a few panoramas sharing the same view of the scene are available. For this matter, we present a matching strategy that augments the accuracy and the number of match points in the context of urban panorama matching. To improve matching results, the method simulates different local transformations at chosen view directions of the panoramas. We show that our matching scheme improves the matching result on the specific panoramas where the classical methods fail to find a sufficient number of matches. This conclusion is supported by real-world experiments performed on 8017 pairs of images coming from 763 different images.

1. Introduction

The capture and processing of a large set of images has many applications. Google Street View and the Photo Explorer of Snavely et al. [13] are two good examples of such applications. The goal of Google Street View is to be able to virtually navigate in large cities following predefined paths. Photo Explorer builds partial 3D representation of cities and buildings from large photo collections allowing unconstrained exploration of scene based on the available multiple views. In this paper, our objective is to match a large set of omnidirectional images sequentially captured in an urban environment.

Different alternative technologies could be used to capture panoramas in a large outdoor or indoor environment. In our case, we use an electric scooter equipped with a computer, GPS and an omnidirectional camera mounted on top.

The resulting panoramic images can then be represented on a 3D surface such as sphere, cylinder or cube [5, 14]. The scooter is driven in an urban environment and panoramas are captured at regular time intervals. Once this is done, the next step is generally to match all these panoramas from which useful 3D pose and structure information can be computed.

Using classical matching strategies on such a set of panoramas is very challenging [2, 11, 12]. These matching schemes are meant to be used on planar images and our experiment has shown that the direct application of these methods on cylindrical panoramas taken in large environments does not produce satisfactory results in many cases. And this for a number of reasons. First, the cylindrical panorama induces non linear distortion of the image. Straight lines that are not vertical do not appear as straight lines on an unfold cylindrical panorama. Therefore, the matching of two different panoramas of the same scene is difficult even with scale and affine invariant features and descriptors such as MSER [8], SIFT [7] or SURF [1].

More importantly, the fact that the images are taken following linear trajectories can produce large changes in the point of view. In addition, urban scenes often include large open spaces (public spaces, parks, parkings) which can limit the number of distinguishing features visible in a given panorama. Figure 1 is a good example of such situation; the images show a single building (the National Gallery of Canada) visible in a limited portion of the panoramas. It becomes therefore essential to maximize the number of good matches at these locations if one wants to connect these panoramas together. Finally, needless to say that, in an outdoor environment, lighting changes, shadows, reflections, and occlusions make difficult obtaining a large number of matches.

Feature detection and matching of non planar images obtained from catadioptric systems was studied in [3, 9]. It was also studied in the case of images with radial distortion [6], wide angle images [4] and spherical panoramas [15]. The strategy used in [3, 9, 15] is to work with re-

projections of the images to be matched. As for [4, 6], they rather modify the convolution step in the definition of the SIFT features so that it applies on wide angles images [4] or on images with radial distortion [6].

One natural solution could be to extract limited fields of view planar images from the panoramas and simply match them all together. But then we need a criterion to select what are the relevant fields of view to extract and what are the relevant images to compare. In [3], regular sampling is used. The approach in [9] creates virtual camera planes at regions defined by MSERs. In our approach, we rather sample the panoramas where the density of SURF features is the highest. This way, we only consider relevant sections of the panoramas for matching. Moreover, even with limited field of view images, large perspective change in the images might prevent the successful matching of two images. This is also the case with planar images and this is one of the reasons why ASIFT [10] was introduced by Morel and Yu. The methods presented in [3, 4, 6, 15] do not take this fact into account. Mauthner et al. [9] uses local affine frame to compensate for perspective changes. These are defined from the convex hull of the MSERs which assumes that these regions can be reliably detected from image to image. Our method consists in extracting limited field of view images extracted at chosen view directions and then matching these images using multiple view reprojections. This approach is inspired by ASIFT with the difference that ASIFT applies multiple global transformations whereas we use multiple local transformations inside the panoramas. Our method can be seen as a generalization and an extension of the simple solution that consists in extracting few precomputed limited fields of view from the panoramas. As in the ASIFT scheme, we aim at identifying the projective images that produces the optimal number of matches through multiple image transformations.

1.1. Multiple Images Matching

One of the most important work in the area of multiple images matching is the one of Snavely et al. [13]. Given a large set of images of the same scene, they explain how to build a 3D representation of the scene in which we can navigate virtually and then collect information about the different objects in the scene. Our problem is different since we follow a long trajectory in which we have only a few panoramas sharing the same view of the scene. This aspect of the problem was also studied in [9, 15] with different techniques.

In this work, our objective was to making sure that the full set of panoramas remains as connected as possible at all geographical locations. For this matter, we developed a matching strategy that augments the accuracy and the number of match points in the context of urban panorama matching. We first describe a simple criterion that enables us to

select pairs of panoramas to be matched. We also present, and this is our main contribution (refer to Subsection 2.4), a matching scheme that augments the matching result by applying the scheme on the specific panorama where the classical methods fail to find a sufficient number of matches. The proposed matching scheme does not replace the classical matching scheme used in large image set such as the one described in [13] or [15], it is rather a supplementary step in the matching phase.

The paper is structured as follows. In Section 2, we present the proposed matching scheme. In Section 3, we present some mathematical background of our matching scheme. Section 4 contains experimental results. These experiments are based on 8017 pairs of images coming from 763 different images. We conclude in Section 5.

2. The Algorithm

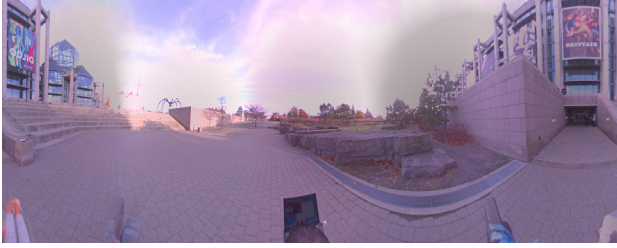
We are given a set of n panoramas P_i ($0 \leq i < n$) that were taken in different outdoor scenes. For each P_i , we have access to the GPS coordinates (x_i, y_i) of P_i . The width and the height of P_i are denoted by *width* and *height* respectively. Since the panoramas were sequentially captured, we know a priori that the predecessors of a panoramas (the ones taken just before) and its successors (the ones taken just after) have been taken at nearby locations (approximately 2m to 4m between each panorama). In addition to comparing a panorama P_i with its predecessors and successors, we also want to compare P_i with all panoramas in the set that are geographically close. Indeed, during a capture session, the trajectories often intersect with each other. In this section, the different steps of the algorithm will be illustrated through the pair of images of Figure 1. These two images are not part of the experimental results presented in Section 4. We use this pair of images in this paper because they intuitively support our explanations.

2.1. Step 1: Find Relevant Pairs of Panoramas to Match

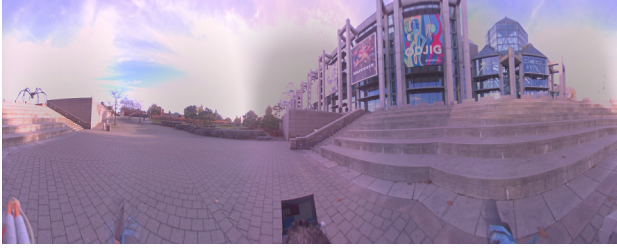
Let L be the list of all pairs of panoramas to be matched. At first, $L = \{ \}$, and at the end of Step 1, L contains the list of all pairs of panoramas to be matched. This first step describes how we compute L . The list of panoramas to be matched is determined from the ℓ predecessors and the ℓ successors of each panorama, if they exist (refer to Figure 2). The 2ℓ neighbouring panoramas of a panorama P_i define a disk D_i centered at (x_i, y_i) . Together with the 2ℓ neighbours of P_i , we also add to L all pairs (P_i, P_j) such that $P_j \in D_i$. This ensures that all panoramas geographically close to P_i are tested for matching since we need to connect as much panoramas as possible.

For each $0 \leq i < n$, do the following.

1. For all $i < j \leq \min\{i + \ell, n\}$, update $L = L \cup$



(a) Image 1 of the National Gallery of Canada.



(b) Image 2 of the National Gallery of Canada.

Figure 1. Two different views of the National Gallery of Canada, Ottawa, Ontario, Canada. For both of them, $width = 1608$ and $height = 640$.

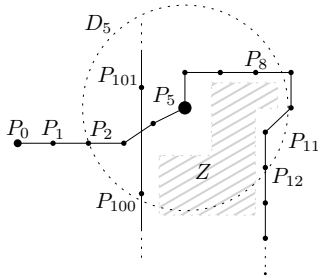


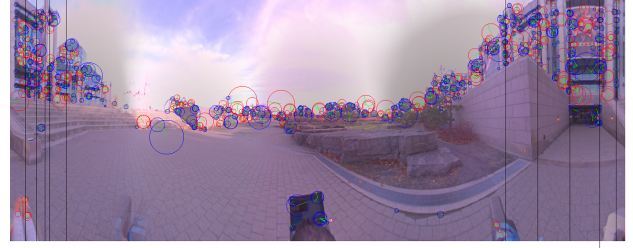
Figure 2. Example of trajectory followed by the scooter. Panoramas are taken at regular intervals. With $\ell = 3$, the disk D_5 includes panoramas $P_2, P_3, P_4, P_5, P_6, P_7, P_8, P_{11}, P_{12}, P_{100}$ and P_{101} . P_{100} and P_{101} come from a later trajectory. The radius of D_5 is $\|P_5 P_2\|$. The grey zone Z can be any kind of object, from a parc without trees to immense buildings. If it is a tall building, then there is no hope to match P_5 with P_{11} nor P_5 with P_{12} .

$\{(P_i, P_j)\}$. (Since we start at $i = 0$, we never need to look at the predecessors of P_i .)

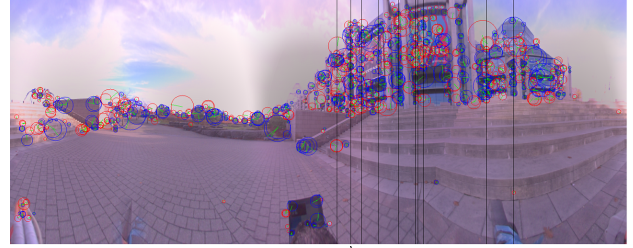
2. Let $r = \max_{\substack{i-\ell \leq j \leq i+\ell \\ 0 \leq j < n}} \|(x_i, y_i) - (x_j, y_j)\|$.
3. For all $j \in [0, n-1] \setminus [i-\ell, i+\ell]$, if $\|(x_i, y_i) - (x_j, y_j)\| \leq r$, then update $L = L \cup \{(P_i, P_j)\}$.

2.2. Step 2: Compute SURF Features

For all $0 \leq i < n$, compute the SURF features of the cylindrical panorama P_i and store them in $F^{(i)}$. In our case, we adjust the threshold of the SURF algorithm such that



(a) Image 1 of the National Gallery of Canada.



(b) Image 2 of the National Gallery of Canada.

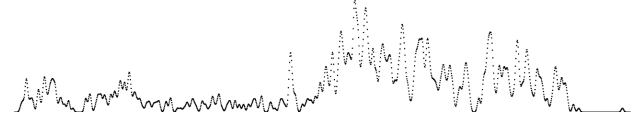
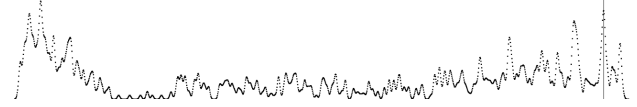
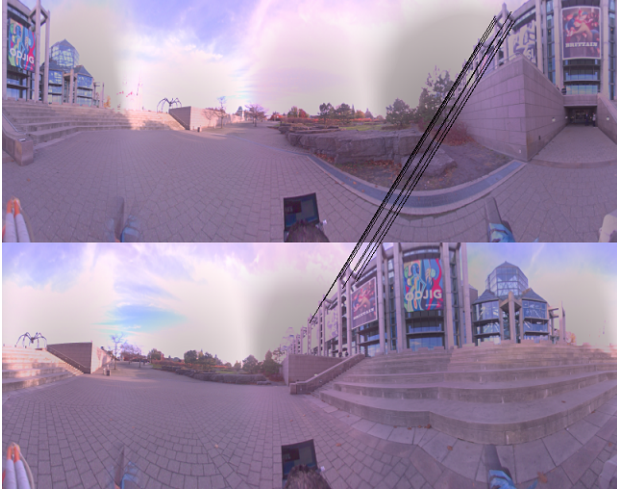


Figure 3. Two different views of the National Gallery of Canada together with their SURF features. The ten vertical lines indicate the ten direction of highest SURF features density (we work with $m = 10$, refer to Subsection 2.4). Under these two images, we show the relative density of SURF features for each direction.

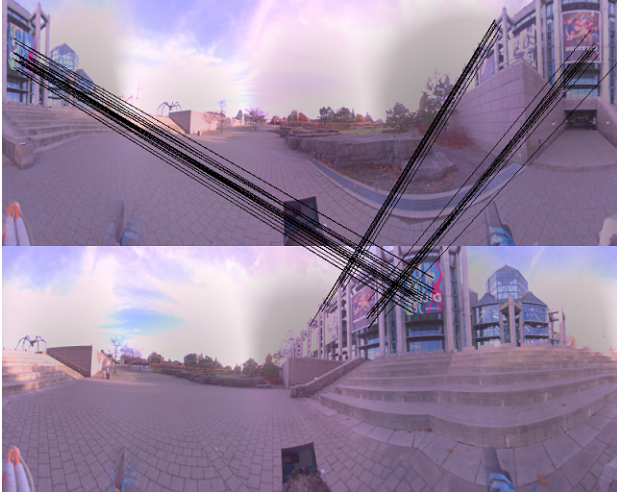
$|F^{(i)}| \approx \frac{1}{1000} width \cdot height$. Figure 3 shows the detected SURF features.

2.3. Step 3: Regular Panorama Matching

For all pairs $(P_i, P_j) \in L$, symmetrically match P_i and P_j using $F^{(i)}$ and $F^{(j)}$ by comparing nearest neighbor with second nearest distance (refer to [7]). Store the matches in $M^{(i,j)}$. Apply RANSAC to $M^{(i,j)}$ and store the matches that support the computed epipolar geometry in $\hat{M}^{(i,j)}$. In the case of our tested pair of panoramas, we obtain 11 matches (refer to Figure 4(a)). We used symmetrical matching and second nearest neighbor constraint to minimize the number of false matches such that RANSAC can successfully identify the underlying epipolar geometry (i.e. the fundamental matrix). Indeed, with difficult pairs of panoramas, we observed that simpler matching scheme gives rise to a very large number of false matches and only few more good matches.



(a) Symmetrical matching with second nearest neighbour constraint (11 matches).



(b) Matching with the tangent plane technique ($m = 9$) (73 matches).

Figure 4. Comparison of the two matching schemes: SURF with second nearest neighbour constraint and the tangent plane technique.

2.4. Step 4: Compute SURF Features on Rotated Tangent Planes

This step contains our main contribution for matching panoramas where the classical methods fail to find enough matches. The mathematical background of this step is described in Section 3.

The proposed strategy is to extract several fields of view images from the panoramas. Given a panorama P_i , the criterion we use is to select the set of directions $\Theta_i = \{\theta_0^{(i)}, \theta_1^{(i)}, \dots, \theta_m^{(i)}\}$ of highest SURF features density. These m directions are rich in SURF features and therefore are worth matching (refer to Figure 3). One way to generate

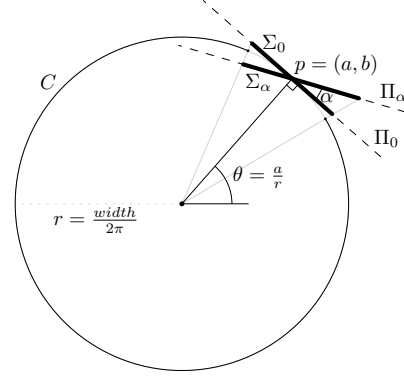


Figure 5. Cylinder C related to Panorama P seen from above. Π_0 is the plane tangent to C at direction θ . Π_α is the plane obtained by rotating Π_0 of an angle α with respect to the axis defined by the tangential line at direction θ . We do not project all of P onto Π_α . We project only $\frac{1}{10}$ of it. This projection defines a screen Σ_α on Π_α . We will then try to match these screens (refer to Subsection 2.5). The grey part of C represents arc of length $\frac{1}{10}width$.

planar fields of view images from P_i is to define a plane $\Pi_0^{\theta_k^{(i)}}$ tangent to P_i at direction $\theta_k^{(i)}$ and then project P_i onto $\Pi_0^{\theta_k^{(i)}}$ (refer to Figure 5). However, we want to take into account the non linear distortion of the image induced by the cylinder that make the matching difficult between two different panoramas of the same scene. Therefore, we also consider the planes $\Pi_\alpha^{\theta_k^{(i)}}$ obtained by rotating $\Pi_0^{\theta_k^{(i)}}$ of an angle α with respect to the axis defined by the tangential line at direction $\theta_k^{(i)}$ (refer to Figure 5). This is inspired by ASIFT in which different possible affine transformations of the images to be matched are generated. Following this idea, not only do we project P_i onto $\Pi_0^{\theta_k^{(i)}}$, we also project P_i onto all $\Pi_\alpha^{\theta_k^{(i)}}$'s. This approach will be referred to as the *tangent plane technique* (TPT). Note that even if this step is time consuming, it is applied only on the restricted set of pairs of panoramas that have been found difficult to match using *regular matching* (REG) (refer to Subsection 2.3).

For all panorama P_i such that there is a panorama P_j with $|\hat{M}^{(i,j)}| < thresh$, do the following.

1. From $F^{(i)}$, compute the set $\Theta_i = \{\theta_0^{(i)}, \theta_1^{(i)}, \dots, \theta_m^{(i)}\}$ of directions with highest vertical SURF features density in P_i .
2. For all $\theta_k^{(i)} \in \Theta_i$, let $\Pi_0^{\theta_k^{(i)}}$ be the plane tangent to P_i at direction $\theta_k^{(i)}$. Let $\Pi_\alpha^{\theta_k^{(i)}}$ be obtained from $\Pi_0^{\theta_k^{(i)}}$ by a rotation of angle α with respect to the axis defined by the tangential line at direction $\theta_k^{(i)}$. Given a restricted field of view of width $\frac{1}{10}width$ centered at direction

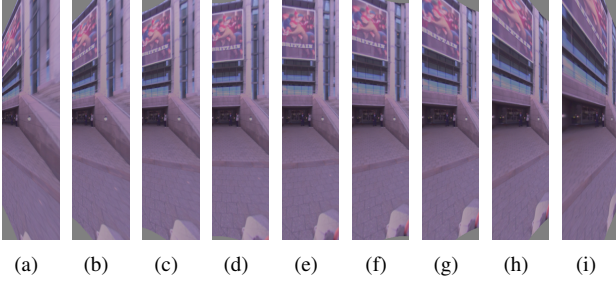


Figure 6. Consider the one before last vertical line in Figure 3(a). It corresponds to the direction of second highest SURF features density. Denote this direction by θ . Let Π_0^θ be the plane tangent to C at direction θ . Let Π_α^θ be obtained from Π_0^θ by a rotation of angle α with respect to the axis defined by the tangential line at direction θ . We project $\frac{1}{10}$ of the panorama on Π_α^θ and we get Σ_α^θ . Figures 6(a)-6(i) show Σ_α^θ for nine different values of α . 6(a) $\alpha = -57^\circ$, 6(b) $\alpha = -43^\circ$, 6(c) $\alpha = -29^\circ$, 6(d) $\alpha = -14^\circ$, 6(e) $\alpha = 0^\circ$ (this angle corresponds to the plane tangent to C at direction θ), 6(f) $\alpha = 14^\circ$, 6(g) $\alpha = 29^\circ$, 6(h) $\alpha = 43^\circ$, 6(i) $\alpha = 57^\circ$

$\theta_0^{(i)}$, project P_i onto $\Pi_\alpha^{\theta_0^{(i)}}$ for all $-\frac{1}{3}\pi \leq \alpha \leq \frac{1}{3}\pi$ with a step¹ of $\frac{1}{125}\pi$. Let $\Sigma_\alpha^{\theta_0^{(i)}}$ be the projection of this restricted field of view onto $\Pi_\alpha^{\theta_0^{(i)}}$. $\Sigma_\alpha^{\theta_0^{(i)}}$ is a rectangle in $\Pi_\alpha^{\theta_0^{(i)}}$.

3. For all $\Sigma_\alpha^{\theta_0^{(i)}}$, compute the SURF features and store the results in $F_{k,\alpha}^{(i)}$ (with the same SURF threshold used for P_i).

The directions with highest vertical SURF features density are depicted in Figure 3. In Figure 6, we see the projection of a panorama onto $\Pi_\alpha^{\theta_0^{(i)}}$ for a fixed direction and for nine different values of α . Figure 6(e) corresponds to the tangent plane $\Pi_0^{\theta_0^{(i)}}$. To understand why this method takes into account the non linear distortion of the image induced by the cylinder, look at Figure 6(c). Compare the ‘‘Britain’’ poster with the one of Figure 1(b). We see the poster from a comparable point of view in Figure 6(c) (that comes from Figure 1(a)) and in Figure 1(b). Several matches obtained by the tangent plane technique come from this part of the panoramas (refer to Figure 4(b)).

2.5. Step 5: Match Panoramas with all SURF Features

Take $(P_i, P_j) \in L$. We use the following strategy based on the directions Θ_i and Θ_j . Each direction $\theta_k^{(i)} \in \Theta_i$ is

¹The range and the number of plane to generate was empirically determined.

matched with all directions $\theta_{k'}^{(j)} \in \Theta_j$ to find the best possible match of directions. For a fixed direction $\theta_k^{(i)}$, we do not compare all $\Sigma_\alpha^{\theta_k^{(i)}}$'s with all $\Sigma_{\alpha'}^{\theta_{k'}^{(j)}}$'s. We rather compare $\Sigma_0^{\theta_k^{(i)}}$'s with all $\Sigma_{\alpha'}^{\theta_{k'}^{(j)}}$'s, and then we compare all $\Sigma_\alpha^{\theta_k^{(i)}}$'s with all $\Sigma_0^{\theta_{k'}^{(j)}}$'s. It is significantly less time consuming this way and experimentation shows that the results are comparable. Once the best possible direction in P_j is found for the direction $\theta_k^{(i)} \in \Theta_i$, the matches between the two corresponding planar images are added to $M^{(i,j)}$.

Let $M^{(i,j)} = \{\}$. For all pairs $(P_i, P_j) \in L$ with $|\hat{M}^{(i,j)}| < thresh$, do the following.

1. For all $\theta_k^{(i)} \in \Theta_i$, do the following.
 - (a) Match all $\Sigma_0^{\theta_k^{(i)}}$'s with all $\Sigma_{\alpha'}^{\theta_{k'}^{(j)}}$'s ($0 \leq k' \leq m$, $-\frac{1}{3}\pi \leq \alpha' \leq \frac{1}{3}\pi$) with second nearest neighbour constraint. Store the results in $M_{(k,0),(k',\alpha')}^{(i,j)}$.
 - (b) Match all $\Sigma_\alpha^{\theta_k^{(i)}}$'s ($-\frac{1}{3}\pi \leq \alpha \leq \frac{1}{3}\pi$) with all $\Sigma_0^{\theta_{k'}^{(j)}}$'s ($0 \leq k' \leq m$) with second nearest neighbour constraint. Store the results in $M_{(k,\alpha),(k',0)}^{(i,j)}$.
 - (c) Take $0 \leq k' \leq m$, $-\frac{1}{3}\pi \leq \alpha \leq \frac{1}{3}\pi$ and $-\frac{1}{3}\pi \leq \alpha' \leq \frac{1}{3}\pi$ such that $|M_{(k,\alpha),(k',\alpha')}^{(i,j)}|$ is maximum. Update $M^{(i,j)} = M^{(i,j)} \cup M_{(k,\alpha),(k',\alpha')}^{(i,j)}$.
2. Apply RANSAC to $M^{(i,j)} \cup M^{(i,j)}$ and store the results in $\hat{M}^{(i,j)}$.

The result of this step is depicted in Figure 4(b).

3. Mathematical Foundation

In this section, we present the mathematics underlying the transformations described in Section 2. We present the results without proofs due to lack of space. Most of the formula presented were derived from 3D analytic geometry formula and rotation matrices.

Let P be a cylindrical panorama. P is a rectangle of size *width* by *height*. Therefore, the radius of the cylinder C it represents is $r = \frac{width}{2\pi}$. Let $p = (a, b)$ be a point on the panorama. On C , p corresponds to the direction $\theta = \frac{a}{r}$ and corresponds to the point $(r \sin(\theta), \frac{1}{2}height - b, -r \cos(\theta))$. Let $\phi(x)$ be the direction corresponding to the abscissa x and, conversly, $\xi(\theta)$ be the abscissa corresponding to the direction θ . We have

$$\begin{aligned} \phi(x) &= \frac{x}{r} \\ \xi(\theta) &= r\theta \end{aligned}$$

Denote by $Cyl(x, y)$ the 3D coordinates on the cylinder corresponding to the point (x, y) on P . Conversely, denote by

$Pano(x, y, z)$ the 2D coordinates on the panorama corresponding to the point (x, y, z) on the cylinder. We have

$$\begin{aligned} & Cyl(x, y) \\ = & \left(r \sin(\theta), \frac{1}{2}height - y, -r \cos(\theta) \right) \\ & Pano(x, y, z) \\ = & \left(\xi(\text{atan2}(x, -z)), \frac{1}{2}height - y \right) . \end{aligned}$$

Therefore, the equation of the plane Π_0 tangent to C at direction $\phi(a)$ is

$$\Pi_0 : z = \frac{x \sin(\phi(a)) - r}{\cos(\phi(a))} .$$

Let Π_α be obtained from Π_0 by a rotation of angle α with respect to the axis defined by the tangential line at direction $\phi(a)$. The equation of Π_α is

$$\Pi_\alpha : z = \frac{x \sin(\phi(a) + \alpha) - r \cos(\alpha)}{\cos(\phi(a) + \alpha)} .$$

We now project C onto Π_α . Let (x, y, z) be a point on the cylinder C . Let Δ be the line through $(0, 0, 0)$ and (x, y, z) . We denote by $TP_\alpha(x, y, z)$ the intersection point of Δ and Π_α . Conversely, Let (x, y, z) be a point on Π_α . Let Δ be the line segment from $(0, 0, 0)$ to (x, y, z) . We denote by $TP_\alpha^{-1}(x, y, z)$ the intersection point of Δ and C . We have

$$\begin{aligned} & TP_\alpha(x, y, z) \\ = & \frac{r \cos(\alpha)}{x \sin(\phi(a) + \alpha) - z \cos(\phi(a) + \alpha)}(x, y, z) \\ & TP_\alpha^{-1}(x, y, z) \\ = & \frac{r}{\sqrt{x^2 + z^2}}(x, y, z) . \end{aligned}$$

When we project C onto Π_α given a restricted field of view of width $\frac{1}{10}width$, we get planar image on a screen Σ_α that is part of Π_α . When $\alpha = 0$, the center of Σ_0 correspond to the direction $\phi(a)$. However, when $\alpha \neq 0$, the center of Π_α does not correspond to $\phi(a)$ (refer to Figure 5). We need the following formulas to take this phenomenon into account and to switch between the two dimensions of the planar image on Σ_α and the three dimensions of the points on Π_α . Given a point (x, y, z) on Π_α , its corresponding coordinates on Σ_α are denoted by $\Pi_\alpha(x, y, z)$. Conversely, given a point (x, y) on Σ_α , its corresponding coordinates on

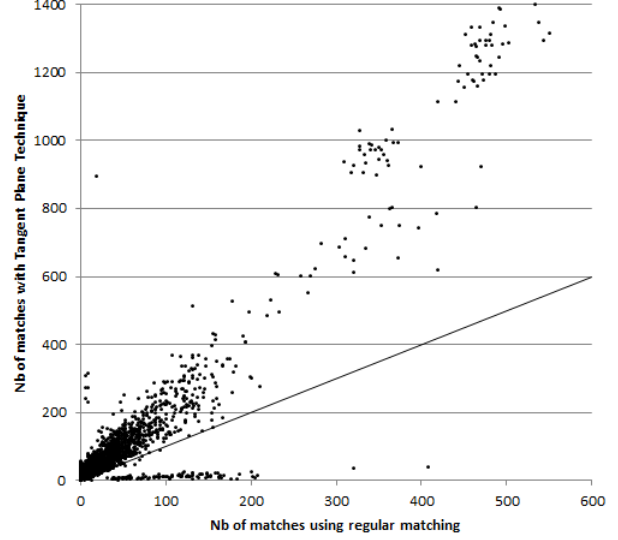


Figure 7. Global matching results of the 8017 pairs of panoramas.

Π_α are denoted by $\Sigma_\alpha^{-1}(x, y, z)$. We have

$$\begin{aligned} & \Sigma_\alpha(x, y, z) \\ = & \left(\frac{r \cos(\alpha) \sin(\phi(a) + \alpha) - x}{\cos(\phi(a) + \alpha)}, \frac{1}{2}height - b - y \right) \\ & \Sigma_\alpha^{-1}(x, y) \\ = & (-x \cos(\phi(a) + \alpha) + r \cos(\alpha) \sin(\phi(a) + \alpha), \\ & \frac{1}{2}height - b - y, \\ & -x \sin(\phi(a) + \alpha) - r \cos(\alpha) \cos(\phi(a) + \alpha)) . \end{aligned}$$

Therefore, the value at pixel (x, y) on the screen Σ_α is equal to the value at pixel $(x', y') = Pano(TP_\alpha^{-1}(\Sigma_\alpha^{-1}(x, y)))$ on P . Since x' and y' are more likely not to be integers, we compute the value of the pixel at position (x, y) by bilinear interpolation.

4. Experimental Results

We tested our method on 8017 pairs of panoramas taken on the campus of the University of Ottawa and in the city of Ottawa. The cylindrical panoramas have an height of 512 pixels and a width of 1608 pixels. The global matching results of the 8017 pairs of panoramas appears in Figure 7. A pair (P_i, P_j) is represented by a point (x, y) on this graph where x is the number of matches between P_i and P_j using regular matching and y is the number of matches obtained using the tangent plane technique for the same pair. The graph shows that we do worst than SURF for 1.9% of the pairs, as good as SURF for 48.7% of the pairs and better than SURF for 49.4% of the pairs.

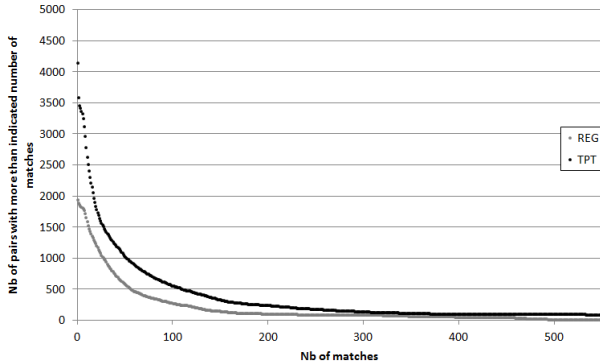


Figure 8. Evolution of the number of pairs of panoramas having a number of matches greater than a given value.

However, we use the tangent plane technique only when the number of matches using regular matching is insufficient. Let us say that a pair of panoramas is considered successfully matched if there are more than 10 good matches. Among the pairs that were not successfully matched using regular matching, 17.1% of them were successfully matched with the tangent plane technique. Among the pairs for which we got between 10 and 20 good matches using regular matching, the average number of matches increases from 14.1 to 35.2 (an increase of 149.2%). Moreover, our method also produces good results on pairs of images for which the regular matching scheme is successful. Indeed, among the pairs for which we got at least 100 matches using regular matching, the average number of matches increases from 224.9 to 468.2 (an increase of 108.2%).

Figure 8 shows the evolution of the number of pairs of panoramas having a number of matches greater than a given value for the two compared approaches (in black, the tangent plane method and in grey, the regular matching scheme). That is a point (x, y) on these curves means that y pairs of panoramas have at least x matches. The fact that the black curve stays over the grey one shows that the tangent plane technique always finds a higher number of pairs of panoramas having at least a given number of matches.

Table 1 gives the number of pairs of panoramas for which the number of matches found is within different intervals. The indicated numbers well emphasize the fact that our large scale panorama matching problem remains very challenging. Indeed, more than half of the pairs of panoramas were not successfully matched. As mentioned in the introduction, many geographically close panoramas suffer from occlusion, difficult lighting condition and low number of reliable features. Nevertheless, The tangent plane technique meets its objective in increasing the success rate of the matching. Figures 9 and 10 illustrate how the tangent plane technique can improve the matching rate.

Nb of good matches	Nb of pairs with REG	Nb of pairs with TPT
0	6080	3884
]0, 10]	354	1508
]10, 20]	377	841
]20, 50]	625	761
]50, 100]	310	470
]100, ∞ [271	553
Total	8017	8017

Table 1. Number of pairs of panoramas for which the number of matches found is within different intervals.



(a) Symmetrical matching with second nearest neighbour constraint (0 match).



(b) Matching with the tangent plane technique ($m = 9$) (25 matches).

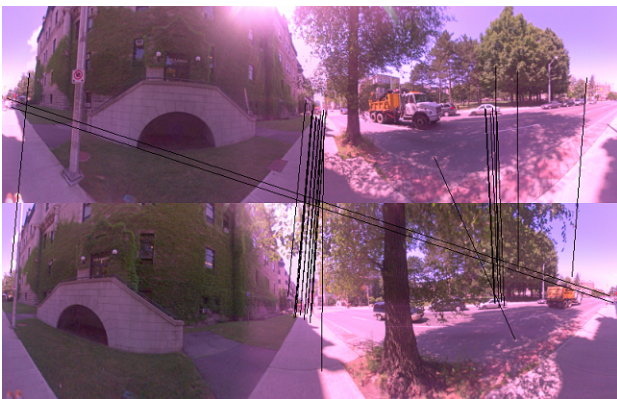
Figure 9. The tangent plane technique can improve the matching rate. Note that the second panorama was not in the list of predecessors/successors of the first panorama.

5. Conclusion

This paper presented a method to increase the number of successfully matched omnidirectional views in the context of large urban image-based models. To demonstrate the validity of such matching scheme, it must be tested on a large number of real-world outdoor images taken under



(a) Symmetrical matching with second nearest neighbour constraint (10 match).



(b) Matching with the tangent plane technique ($m = 9$) (38 matches).

Figure 10. The tangent plane technique can improve the matching rate. In this example, the second panorama was the second successor of the first panorama.

various viewing conditions. Here, we used 8017 city panorama pairs on which we showed an increase of 100. The method is based on multiple planar reprojections of local virtual planes. These planes are defined on tangential lines of the cylindrical panoramas taking advantage of the fact that these omnidirectional images were captures from translations on a ground plane. Generalization of this matching technique to any kind of panoramic representation (e.g. spherical) is also possible using the notion of tangent plane. Using this approach, it takes a few minutes to match a pair of panoramas. This one is therefore, meant to be used offline, on those panoramas where regular matching schemes fail to find an adequate number of matches. The next step will then be to build long tracks of matched points across multiple panoramas, following an approach similar to the one described in [13] and to perform 3D pose and structure estimation.

Acknowledgements: The authors thank Isabelle Michaud

for helpful discussion.

References

- [1] H. Bay, A. Ess, T. Tuytelaars, and L. J. V. Gool. Speeded-up robust features (surf). *Computer Vision and Image Understanding*, 110(3):346–359, 2008. 1
- [2] M. Brown and D. G. Lowe. Automatic panoramic image stitching using invariant features. *International Journal of Computer Vision*, 74(1):59–73, 2007. 1
- [3] C. Demonceaux and P. Vasseur. Omnidirectional image processing using geodesic metric. In *ICIP*, pages 221–224, 2009. 1, 2
- [4] P. Hansen, P. Corke, W. W. Boles, and K. Daniilidis. Scale invariant feature matching with wide angle images. In *IROS*, pages 1689–1694, 2007. 1, 2
- [5] F. Kangni and R. Laganière. Epipolar geometry for the rectification of cubic panoramas. In *CRV*, page 70, 2006. 1
- [6] M. Lourenço, J. P. Barreto, and A. Malti. Feature detection and matching in images with radial distortion. In *ICRA*, pages 1028–1034, 2010. 1, 2
- [7] D. G. Lowe. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60(2):91–110, 2004. 1, 3
- [8] J. Matas, O. Chum, M. Urban, and T. Pajdla. Robust wide baseline stereo from maximally stable extremal regions. In *BMVC*, 2002. 1
- [9] T. Mauthner, F. Fraundorfer, and H. Bischof. Region matching for omnidirectional images using virtual camera planes. In *Proceedings of the 11th Computer Vision Winter Workshop*, pages 93–98, 2006. 1, 2
- [10] J.-M. Morel and G. Yu. Asift: A new framework for fully affine invariant image comparison. *SIAM J. Imaging Sciences*, 2(2):438–469, 2009. 2
- [11] G. Roth. Automatic correspondences for photogrammetric model building. In *20th Congress of ISPRS - International Society for Photogrammetry and Remote Sensing*, pages 713–720, 2004. 1
- [12] I. Skrypnik and D. G. Lowe. Scene modelling, recognition and tracking with invariant image features. In *ISMAR*, pages 110–119, 2004. 1
- [13] N. Snavely, S. M. Seitz, and R. Szeliski. Photo tourism: exploring photo collections in 3d. *ACM Trans. Graph.*, 25(3):835–846, 2006. 1, 2, 8
- [14] R. Szeliski and H.-Y. Shum. Creating full view panoramic image mosaics and environment maps. In *SIGGRAPH*, pages 251–258, 1997. 1
- [15] A. Torii, M. Havlena, and T. Pajdla. From google street view to 3d city models. In *Computer Vision Workshops (ICCV Workshops), 2009 IEEE 12th International Conference on*, pages 2188–2195, 27 2009-oct. 4 2009. 1, 2