Cylindrical panoramic transfer and appearance prediction for the match of real observations

Livatino, Salvatore

Published in:
CVonline : the Evolving, Distributed, Non-Proprietary, On-Line Compendium of Computer Vision

Publication date:
2003

Document Version
Publisher's PDF, also known as Version of record

Cylindrical Panoramic Transfer and Appearance Prediction for the Match of Real Observations

Salvatore Livatino

Email: sl@cvmt.dk

The summation of natural phenomena which can be synthesized in realistically generated virtual views may provide the user with a strong sense of presence (based on the visual realism). This can be the case, even if not all the original physical effects are "correctly" transferred to the newly generated view. A physically-based image mapping is consequently not always necessary in order to provide visual realism. However, when the goal of virtual-view synthesis is not a realistic visualization but, as in case of (Livatino [14]), the match of real observations, a physically based image mapping represents an important factor to take into account, and it has a great influence on the way the image-matching process should be designed, (e.g. if it should be feature- or correlation- based).

The class of image-based rendering techniques characterized as *"geometric-valid pixel reconstruction"*, [11], typically uses relatively small number of images because of the application of geometric constraints, (either recovered at some stage or known a priori), to reproject image pixels appropriately at a given camera viewpoint. The geometric constraints can be of the form of known depths or correspondence values, epipolar constraints between pairs of images, or trilinear tensors that link correspondences between triplets of images. The geometric constraints can also be exploited to solve the visibility problem, (i.e. when an object or scene surface appears in front of another object or surface, even if it should lie behind).

In the literature different possibilities can be found within geometrically-valid pixel reprojection, depending on the chosen method and available information. These are mainly characterized by the use of trilinear tensors, (Shashua [21], Avidan et al. [1], Hartley [10]) and fundamental matrix, (Faugeras [9], Leveau et al. [13]). The image reprojection is very often based on a direct exploitation of known depths, correspondences, (Chen and Williams [6], Chang et al. [5], Seitz and Dyer [20]), epipolar constraints, (McMillan and Bishop [19], Kang and Szeliski [12]), etc.

It is in the exploitation of epipolar constraints, that the author have focused his attention and got inspired concerning the transfer of pixel values between cylindrical reference-views to a new viewing position, where the application context is a realistic visualization of landmark visual predictions for the purpose of mobile robot navigation, (Livatino [14], [17]). In particular, based on recent developments in image-based rendering involving the use of the Plenoptic function, (which describes light rays visible at any point in space), and on cylindrical panoramic images, (McMillan and Bishop [19], Kang [11]), it is proposed the *interpolation of cylindrical panoramic images*. The exploitation of cylindrical panoramic

views for the purpose of mobile robot navigation is receiving increasing attention in the recent years, (Yuen and MacDonald [22]). In addition, some authors have very recently announced as their future research activity in mobile robotics, the synthesis of virtual views based on cylindrical panoramic references (Bunschoten and Kröse [4]).

Cylindrical panoramic images can "naturally" be acquired by a robotic system during its navigation in a learning phase, and they can represent the basis from where landmark reference-views can be extracted, (Livatino [14], [15]). Then, during the self-localization phase, the proposed interpolation of cylindrical reference-views can be used to render realistic visual predictions and to match current observations. Figure 1 visually describes the transfer of texture values from cylindrical reference-views to the visual prediction for an example landmark.

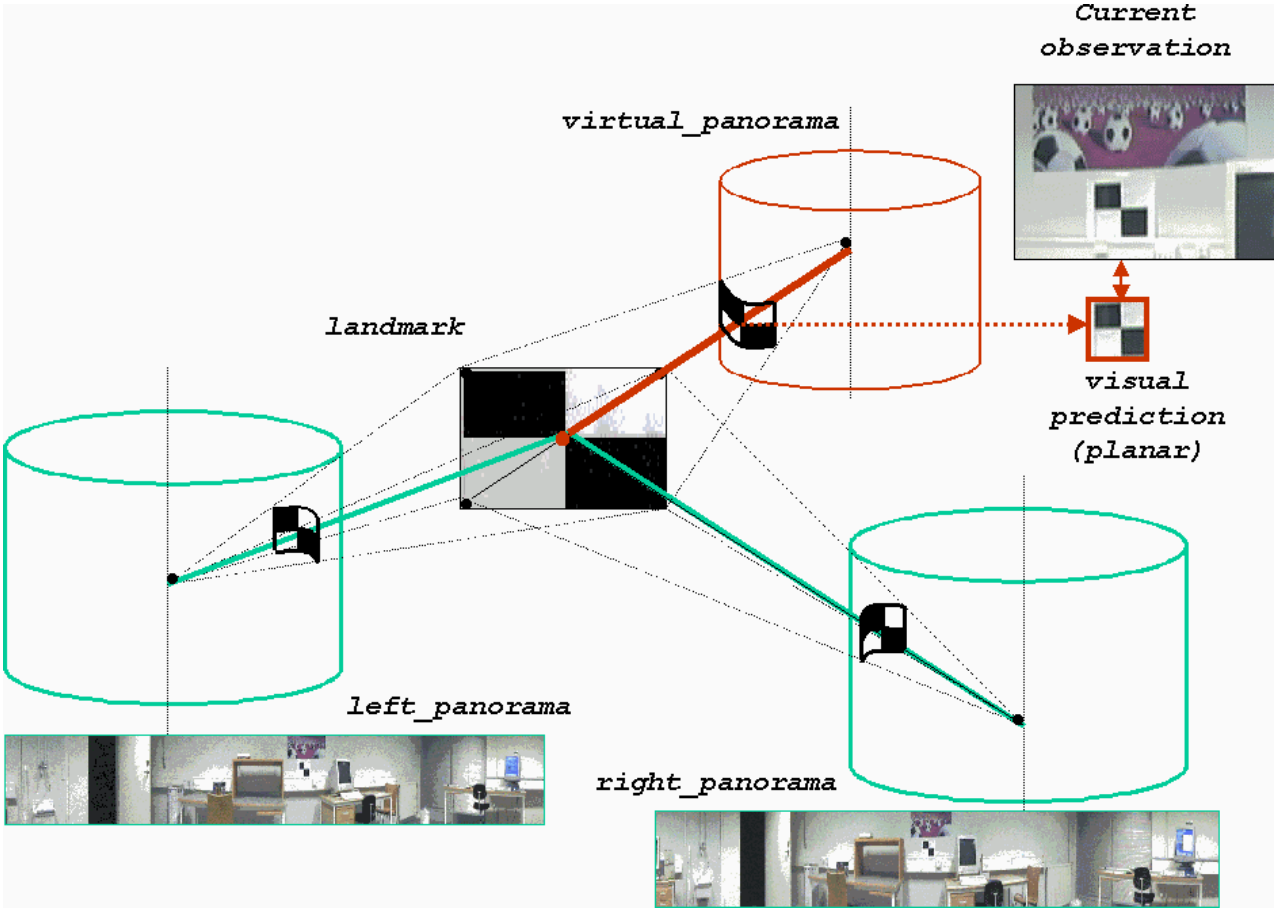

Figure 1: The figure visually describes the transfer of texture values from cylindrical reference-views to the visual prediction for an example landmark, (the "ideal" landmark). The top-right image represents the *"current observation"* where the *visual prediction* represent the view re-mapped from reference cylindrical images according to the current estimate of camera viewpoint. The visual prediction can then used as template in an *image correlation* process to locate the landmark precisely in the image-plane.

The transfer of pixels to the new (virtual) view by interpolation of cylindrical landmark reference-views is proposed to be performed in the following way.

1. *Cylindircal Pixel Transfer.* I.e. mapping geometrical pixel correspondences between reference landmark views and visual prediction. For every pixel in the visual prediction, the corresponding pixels in the reference views are calculated by estimating the angular disparity. This mapping follows the idea of "plenoptic transfer", (McMillan and Bishop [19]).

2. *View-Adapted Texture Mapping.* I.e. mapping texture correspondences between reference views and visual prediction. For every pixel in the visual prediction, its texture value (light intensity) is calculated by a weighted average of corresponding texture values in the reference views. This follows the idea of view-dependent texture mapping, (Debevec et al. [7]), where texture in the visual prediction can primarily be based on reference-images taken from viewpoints which are the closer to current observation. Closer images are in fact expected to best approximate landmark visible aspects and local illumination effects present in current camera observation.

It is based on the realism of virtual-views generated by the proposed method the author aim of improving the match between landmark visual prediction and current observation, (the match between landmark visual prediction and current observation can be used for mobile robot self-positioning, Livatino and Madsen [16]). The proposed technique thus represents the system's "answer" to the issue of synthesizing realistic and geometric-valid visual predictions.

1 Cylindrical Pixel Transfer

The goal of cylindrical pixel transfer is an automatic and reliable estimate of pixel correspondences between reference-views and visual prediction. The available knowledge consists of camera focal length in pixels, landmark position, orientation and reference-views, and an estimate of current robot pose. In addition, there is the knowledge that learned landmarks lie on planar or "almost planar" surfaces.

The basic concept for interpolating cylindrical panoramic images has been shown in figure 1. This is equivalent to computing 3D points from image correspondences and projecting them to a new target image. McMillan and Bishop, [19], devised an efficient method for transferring known image disparity values between cylindrical panoramic images to a new virtual view. Their approach uses the *angular disparity* (related to each cylindrical pair) to automatically generate warps that map reference views to arbitrary cylindrical or planar views.

The angular disparity can be estimated in different ways depending on the available knowledge. For example, by manually or automatically specifying a sparse set of corresponding points that are visible in both reference-views, by knowing or recovering camera internal parameters, and by exploiting epipolar geometry, a dense set of corresponding points can be recovered. Examples of procedures which exploit epipolar relations to recover dense correspondences from a sparse set of corresponding points, can be found in different literature works, (McMillan and Bishop [19], Faugeras [9], Blanc, Livatino and Mohr [3], [2]).

In the case of [14], the angular disparity is inferred from previously estimated correspondences along cylindrical epipolar lines. These correspondences allowed for estimating the geometry of the landmark surface based on: 3D positions of landmark center, a minimum of two landmark corners, and the result of the planarity test, Landmarks are required to lie on one planar or "almost planar" surface, nevertheless, the same procedure could be applied to landmarks lying on more than one surface (in case the surfaces are known).

The proposed rendering system takes as input cylindrical reference-views of landmarks, along with the map of the angular disparities. This information is used to automatically generate image warps that map landmark reference-views to arbitrary cylindrical landmark-views. Note that the generated warps are capable of describing perspective effects, and occlusions (using a simple visibility algorithm that guarantee back-to-front ordering [19]).

The cylindrical-to-cylindrical mapping is illustrated in figure 2. Each angular disparity value, $\Delta_{\gamma,v}$, can be obtained as in equation 1. Note that (γ, v) are the pixel coordinates in the panorama, where γ is an angle while v is the pixel row.

$$\Delta_{(W_\gamma^{left}, W_v^{left})} = W_\gamma^{right} - W_\gamma^{left} \tag{1}$$

where $(W_\gamma^{left}, W_v^{left})$ represents a generic pixel in one of the left cylindrical reference-views (labeled "left") identified by the angle γ and the ordinate v, and $(W_\gamma^{right}, W_v^{right})$ represents the correspondent ordinate v for a certain angle γ, in the right reference-view.

Knowing the angular disparity for each landmark pixel, this can be converted for each position on the left cylinder (γ, v), into an image flow vector field, $(\gamma + \Delta_{\gamma,v}, v(\gamma + \Delta_{\gamma,v}))$. The figure 2 top row illustrates this conversion.

The disparity values can then be transfered from the known cylindrical pairs $(C_x^{left}, C_y^{left}, C_z^{left})$ and $(C_x^{right}, C_y^{right}, C_z^{right})$ (which respectively represent the left and the right camera positions), to a new cylindrical projection in an arbitrary position, $(C_x^{virt}, C_y^{virt}, C_z^{virt})$, using the following equations, where τ is the rotation offset which aligns the angular orientation of the cylinders to a common frame,

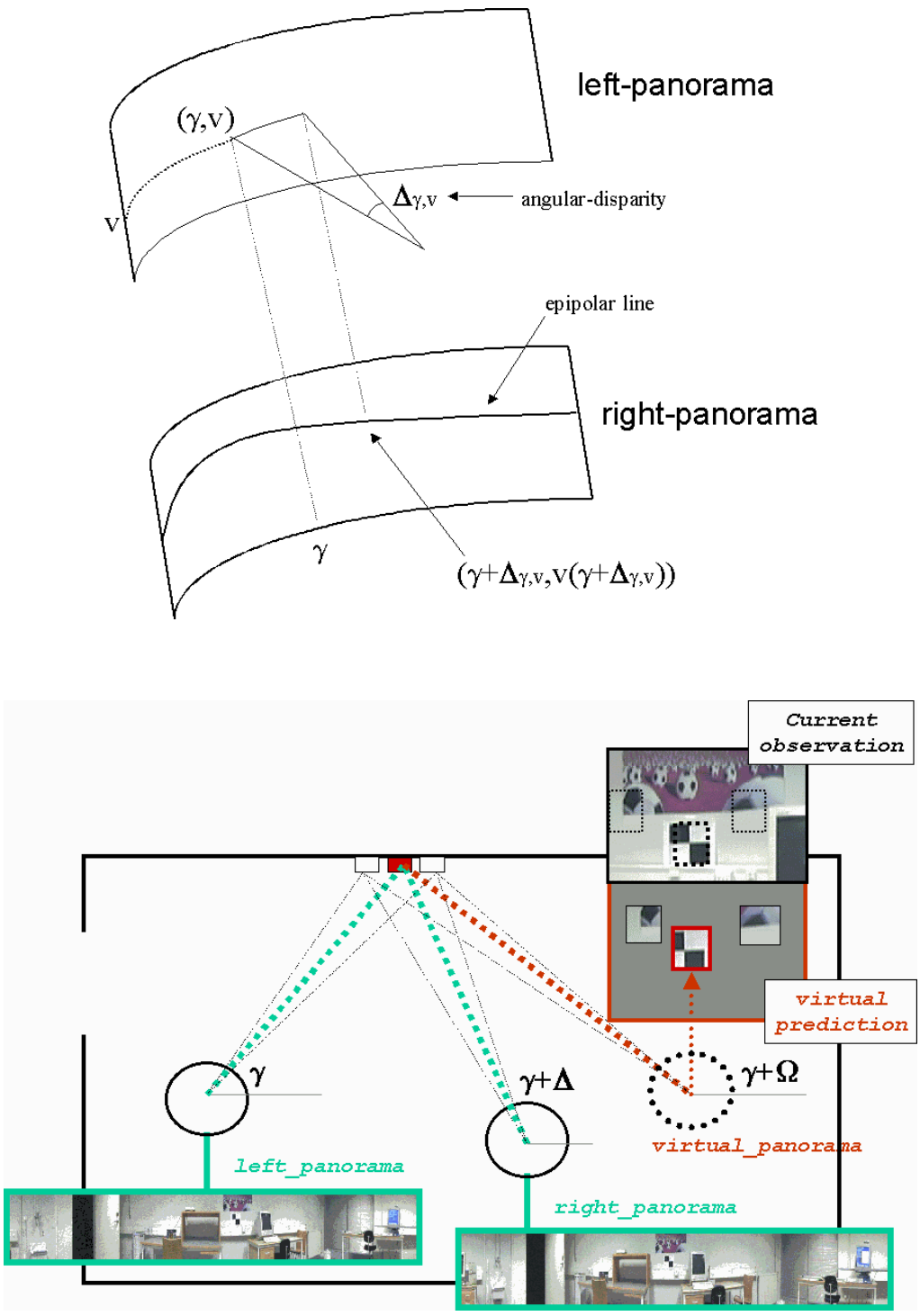

Figure 2: The top figure illustrates pixel correspondences in two different cylindrical panoramas and the related *angular disparity*. The bottom figure illustrates the cylindrical-to-cylindrical mapping based on angular disparity through an example (which consider a workspace floor-map). The bottom figure also includes images of reference panoramas, visual predictions and current observation.

$$
\begin{aligned}
a &= (C_x^{right} - C_x^{virt}) \cos(\tau - W_\gamma^{left}) + (C_y^{right} - C_y^{virt}) \sin(\tau - W_\gamma^{left}) \\
b &= (C_y^{right} - C_y^{left}) \cos(\tau - W_\gamma^{left}) + (C_x^{right} - C_x^{left}) \sin(\tau - W_\gamma^{left}) \\
c &= (C_y^{virt} - C_y^{left}) \cos(\tau - W_\gamma^{left}) + (C_x^{virt} - C_x^{left}) \sin(\tau - W_\gamma^{left})
\end{aligned} \tag{2}
$$

$$
\cot \Omega_{(W_\gamma^{virt}, W_v^{virt})} = \frac{a + b \cot \Delta_{(W_\gamma^{left}, W_v^{left})}}{c} \tag{3}
$$

The resulting $\Omega_{(W_\gamma^{virt}, W_v^{virt})}$ is the angular disparity between the generic pixel in the left cylindrical reference-view, $W_{\gamma,v}^{left}$, and the corresponding pixel in the virtual cylindrical view. In this way, each resulting angular disparity value, $\Omega_{\gamma,v}$, can be converted, for each position on the left cylinder (γ, v), into an image flow vector field ($\gamma + \Omega_{\gamma,v}$, $v(\gamma + \Omega_{\gamma,v})$) using the epipolar relation given by equation 4.

$$
W_v^{virt}(W_\gamma^{virt}) = \frac{M_x \cos(\tau - W_\gamma^{virt}) + M_y \sin(\tau - W_\gamma^{virt})}{M_z} + C_v \tag{4}
$$

where

$$
\begin{bmatrix} M_x \\ M_y \\ M_z \end{bmatrix} = \left(\begin{bmatrix} C_x^{left} \\ C_y^{left} \\ C_z^{left} \end{bmatrix} - \begin{bmatrix} C_x^{virt} \\ C_y^{virt} \\ C_z^{virt} \end{bmatrix} \right) \times \begin{bmatrix} \cos(\tau - W_\gamma^{left}) \\ \sin(\tau - W_\gamma^{left}) \\ C_v - W_v^{left} \end{bmatrix} \tag{5}
$$

and

$$
W_\gamma^{virt} = W_\gamma^{left} + \Omega_{\gamma,v} \tag{6}
$$

where τ is the rotation offset which aligns the angular orientation of the cylinders to a common frame, and C_v is ordinate v of the scan-line where the center of the projection would project onto the scene, (i.e. the ordinate of the line of zero elevation).

The above equation gives a concise expression for the curve, $W_v^{virt}(W_\gamma^{virt})$, (i.e. the cylindrical epipolar line), formed by the projection of a ray across the surface of a cylinder, (labeled "virt"), where the ray is specified by its positions on some other cylinder, (labeled "left").

Once the angular disparity, $\Delta_{\gamma,v}$, has been used for the transfer of the disparity values between the reference cylinder to a new viewing position, each estimated pixel in the virtual cylinder, $W_{i,j}^{virt}$, is projected on the virtual camera image-plane, so becoming $W_{s,t}^{virt-plan}$, in order to generate the landmark visual prediction. The visual prediction is converted to planar to be compared to landmark current observation. Figure 3 illustrates the mapping from cylindrical to planar image.

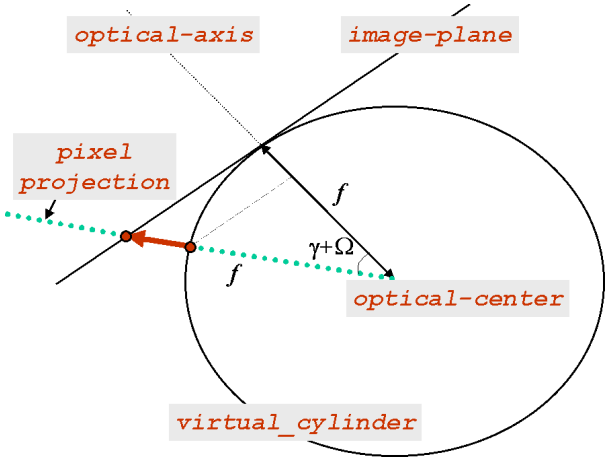

Figure 3: The figure illustrates the mapping from cylindrical to planar image.

The proposed landmark pixel transfer procedure starts with a forward mapping (from reference to virtual) for what concern the landmark corners (which position has been previously established by stereo-matching). This results in a region of the virtual-cylinder delimited by the four projected corners. Each pixel included in the delimited region is then projected forward or inverse depending on the required performance. Having in this case priority a high texture fidelity, a forward mapping is adopted in the compression case and an inverse mapping in the enlargement case.

In summary, the proposed technique of cylindrical pixel transfer allows for establishing pixel correspondence between landmark reference-views and virtual prediction. In particular, two reference views were considered. Experiments, (Livatino [14]), showed that the proposed technique allows for reliable matches between prediction and observation even in presence of significant positional errors, (denoted by clear displacements in image-plane, between prediction and observation).

2 View-Adapted Texture Mapping

Once a correspondence between landmark reference-views and visual prediction has been established, the texture values in the reference-views can be mapped to the virtual view. It is proposed to apply projective texture mapping in an *adaptive* way, in order to take

advantage of multiple reference-views of landmarks, (from different positions), depending on the current robot pose.

The basic concept is that the appearance of an object in the camera image-plane strongly depends on relative position between camera and object, and on present light conditions. In particular, different views of an object may reveal different visible aspects, (e.g. disocculsions), and local illumination effects, (e.g. shadows, reflections, highlights, etc.). Figure 4 summarizes the main factors affecting appearance of objects when observed from different viewpoints.

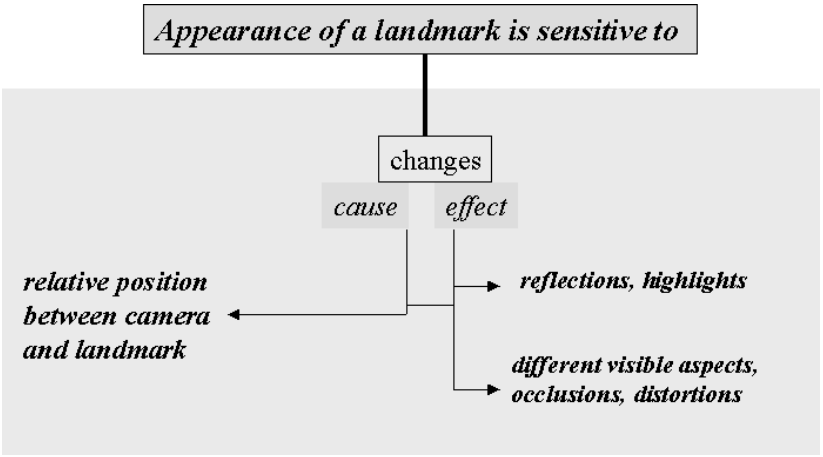

Figure 4: The figure summarizes main factors affecting appearance of objects when observed from different viewpoints.

But, how should texture values of different reference-views contribute when transfered to the same pixel in the virtual view?

In some of the literature works related to realistic visualization the issue of blending multiple images have been addressed, and it has been demonstrated the advantage of considering more than one reference texture when generating texture on a virtual view.

Different techniques have been proposed for blending texture values relative to different views. Among them, simple weighting functions based on the angle of the camera to the object, to more sophisticated post-rendering calculations, (Mark et al [18]). In case a geometric model is available for the represented objects, (even if this is a coarse model), textures could efficiently be mapped by a view-dependent projective mapping as shown in Debevec, Yu and Borshukov [8]. In Debevec, Taylor and Malik [7] it is shown that such a mapping could also be exploited to refine the geometric model of an object by a technique named: *model-based stereo.*

In case of Livatino [14], it is proposed to merge two o more landmark reference views into a composite rendering, combining texture values of correspondent pixels in different reference views. In particular, the system calculates a weighted average where involved

textures provide a different contribution. Images from reference views which are closer to current viewpoint are expected to better approximate the current view than a reference-view further away. Closer images are in fact expected to best approximate landmark visible aspects and local illumination effects present in current camera observation.

Figure 6 shows an example situation which can also be referred to the landmark of figure 5 (a portion of the computer monitor). In case of a planar object with a planar neighbor region, it is not that relevant which reference view should provide a higher contribution when estimating the texture of the current view. In case of an "almost" planar object with a not planar neighbor region (as the case of the monitor), the closest reference view (ref_1 in figure 5) should provide the higher contribution. In fact, the closest view contains reflections, visible aspects, etc., which could have not been shown in the farer reference view (ref_2).

In particular, it is proposed to calculate a weighted average where involved textures provide a different contribution which depends on:

- *the magnitude of the angle* between the lines that connect the landmark center with the optical centers of the camera, (related to considered reference and current view). This angle is depicted in figure 5 as α_i , $i = 1, 2$. The weight for this angle is inversely proportional to the magnitude of the angle.
- *the distance* between the current viewing position and the reference positions. This distance is depicted in figure 5 as d_i , $i = 1, 2$. The weight for this distance is inversely proportional to the length.

The reference view which is closer to current viewpoint in "angle" and in "distance" will thus give a higher contribution in the final summation. In case of two reference views, the resulting texture value for a pixel, VP, would then be calculated as in the following:

$$VP_{\alpha} = Ref_1 \ \frac{\alpha_2}{\alpha_1 + \alpha_2} + Ref_2 \ \frac{\alpha_1}{\alpha_1 + \alpha_2} \tag{7}$$

$$VP_{dist} = Ref_1 \ \frac{d_2}{d_1 + d_2} + Ref_2 \ \frac{d_1}{d_1 + d_2} \tag{8}$$

$$VP = \frac{VP_{\alpha} + VP_{dist}}{2} \tag{9}$$

The above equations can naturally be extended to the case of more than two reference views.

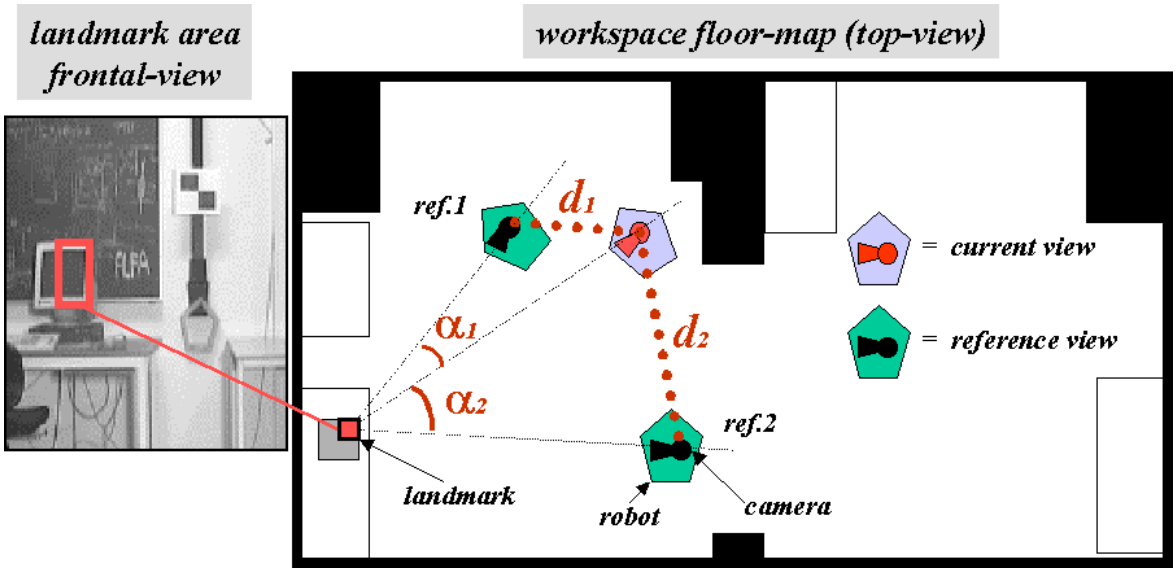

Figure 5: The figure left hand side shows the considered landmark (representing a part of the computer monitor). The figure right-hand side represents the angles between the lines that through the landmark center intersect the camera optical-center (angles α_1 and α_2), and the distances between current viewing position and the reference positions.

Merging reference-views based only on the above criteria can cause visible seams in the landmark visual prediction due to specularity, and unmodeled geometric detail may arise when neighboring textures comes from different reference-images and in case of occlusions or "disocclusions". Some of the techniques proposed in the literature for calculating texture transitions between different mapped views could then be applied to cope with the problem.

In the context of robot navigation as proposed in (Livatino [14]), the main reasons for proposing view-adapted texture mapping can be summarized in:

1. a more reliable match between visual prediction and current observation;

2. an advantageous way of blending landmark reference-views when current viewpoint encompasses an area not entirely observed in one of the reference-views.

References

[1] S. Avidan and A. Shashua. Novel view synthesis in tensor space. In *Conference on computer vision and pattern recognition*, pages 1034–1040, San Juan, Puerto Rico, June 1997.

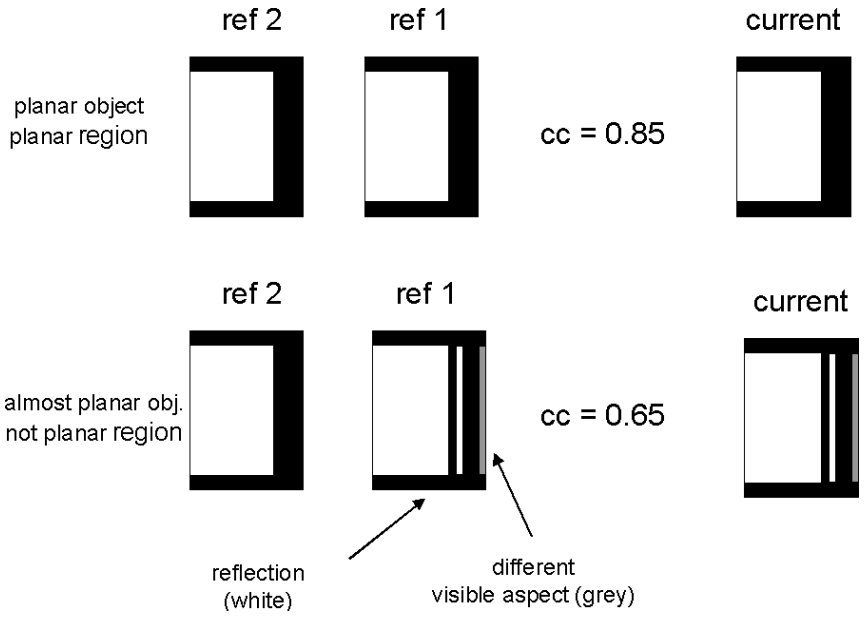

Figure 6: Figure shows an example situation where the reference view ref_1, which is the closest to current view, is expected to provide a better texture information than the reference view ref_2 which is farer. Possible reflections, highlights, shadows, and visible aspects, contained in the closest reference view, best approximate the current view texture. The figures top-row represent a portion of a planar object with a planar neighbor region, while the figures bottom-row represent a portion of an "almost" planar object, (e.g. the computer monitor), with a not planar neighbor region. "cc" represents an example of correlation coefficient which may result when matching the ref_1 and ref_2 views.

[2] J. Blanc, S. Livatino, and R. Mohr. Fast and realistic image synthesis for telemanipulation purposes. In *European Workshop on Hazardous Robotics*, pages 77–83, Barcelona, Spain, November 1996.

[3] J. Blanc and R. Mohr. From image sequence to virtual reality. In E.P. Baltsavias, editor, *ISPRS Intercommission Workshop: From Pixels to Sequences*, pages 144–149, Zurich, Switzerland, March 1995.

[4] R. Bunschoten and B. Kröse. 3-d reconstruction from cylindrical panoramic images. In *9th International Symposium on Intelligent Robotics Systems*, Touslouse, France, July 2001.

[5] N.L. Chang and A. Zakhor. View generation for three-dimensional scenes from video sequences. *IEEE Transaction on Image Processing*, 6(4):584–598, April 1997.

[6] S. Chen and L. Williams. View interpolation for image synthesis. In *Computer Graphics (SIGGRAPH'93)*, pages 279–288, August 1993.

[7] P.E. Debevec, C.J. Taylor, and J. Malik. Modeling and rendering architecture from photographs: A hybrid geometry- and image-based approach. In *Computer Graphics (SIGGRAPH'96)*, pages 11–20, August 1996.

[8] P.E. Debevec, Y. Yu, and G. Borshukov. Efficient view-dependent image-based rendering with projective texture mapping. In *9th Eurographics Workshop on Rendering*, pages 105–116, 1998.

[9] O. Faugeras. Three-dimensional computer vision: a geometric viewpoint. *MIT Press, Cambridge, Massachusetts*, 1993.

[10] R. Hartley. In defence of the 8-point algorithm. In *Fifth international conference on computer vision (ICCV'95)*, pages 1064–1070, June 1995.

[11] S.B. Kang. A survey of image based rendering techniques. *VideoMetrics, SPIE*, 3641:2–16, 1999.

[12] S.B. Kang and R. Szeliski. 3-d scene data recovery using omnidirectional multibaseline stereo. In *IEEE Computer society conference on computer vision and pattern recognition*, pages 364–370, June 1996.

[13] S. Leveau and O. Faugeras. 3-d scene representation as as collection of images and fundamental matrix. Technical Report 2205, INRIA Sophia-Antipolis, February 1994.

[14] S. Livatino. *Acquisition and Recognition of Natural Landmarks for Vision-based Autonomous Robot Navigation.* PhD thesis, Computer Vision and Media Technology Laboratory, Aalborg University, Denmark, June 2003. http://www.cvmt.dk/~sl/phd.

[15] S. Livatino and C.B. Madsen. Autonomous robot navigation with automatic learning of visual landmarks. In *7th International Symposium on Intelligent Robotics Systems (SIRS)*, pages 419–428, Coimbra, Portugal, July 1999.

[16] S. Livatino and C.B. Madsen. Optimization of robot self-localization accuracy by automatic visual-landmark selection. In *The 11th Scandinavian Conference on Image Analysis (SCIA)*, pages 501–506, Kangerlussuaq, Greenland, June 1999.

[17] S. Livatino and C.B. Madsen. Acquisition and recognition of visual landmarks for autonomous robot navigation. In *8th International Symposium on Intelligent Robotics Systems (SIRS)*, pages 269–279, Reading, United Kingdom, July 2000.

[18] W.R. Mark, L. McMillan, and G. Bishop. Post-rendering 3d warping. In *Symposium on interactive 3D graphics*, pages 7–16. ACM Press, April 1997.

[19] L. McMillan and G. Bishop. Plenoptic modeling: an image-based rendering system. In *Computer Graphics (SIGGRAPH'95)*, pages 39–46, August 1995.

[20] S.M. Seitz and C.R. Dyer. View morphing. In *Computer Graphics (SIGGRAPH'96)*, pages 21–30, August 1996.

[21] A. Shashua. Algebraic functions for recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 17(8):779–789, 1995.

[22] D.C.K. Yuen and B.A. MacDonald. Considerations for the mobile robot implementation of panoramic stereo vision system with a single optical centre. In *Image and Vision Computing*, pages 335–339, Auckland, New Zealand, Nov. 2002.