

A New Camera Projection Model and its Application in Reprojection

Kuntal Sengupta

Ivan Poupyrev

Tatsumi Sakaguchi

Jun Ohya

ATR Media Integration & Communications Research Laboratories

Abstract: In this paper we present a new camera projection model, which is intermediate between the affine camera model and the pin hole projection model. It is modeled as a perspective projection of 3D points into an arbitrary plane, followed by an affine transform of these projected points. We observe that the reprojection of a point into a novel image can be achieved uniquely provided that we have located a set of five reference points over four images (of which three are input images, and the fourth is the novel image). Also, the reprojection theory does not assume that the input images are captured from cameras with identical internal calibration parameters. Thus, we apply our technique two different domain: (a) generation of novel images from a stereo pair (b) generation of virtual walkthrough sequence with a monocular image sequence as input.

1 Introduction

Recently, communication between humans located at distant sites has increased in importance. However, existing visual communication means such as video conferencing systems have limitations. For a user, the feeling of situated at distant locations is often difficult to overcome. One solution is to create an environment in which humans can feel that they are co-located in one real space. To realize this, it is necessary to generate views from arbitrary points in the real space. Also, in virtual walkthrough applications where the 3D model of the world is not known, we need to generate an image sequence for the “virtual” walk starting from an input monocular image sequence of the world. In this paper, we introduce a new camera projection model, and develop an affine coordinate based reprojection scheme for image based rendering for both the domains mentioned above.

There are approaches in the literature [1, 4] which first solves the problem of 3D reconstruction either by using the camera calibration parameters, or by assuming that the objects undergo 3D affine or projective transformation. This is followed by the projection of the 3D points into the new image. The process of 3D reconstruction is unstable under errors in image measurements, and a variety of other factors. Direct approaches for scene reprojection and novel view synthesis, such as epipolar line intersection, either require strong calibration of the cameras, or at least eight matched points in three views [2].

A New Camera Projection Model and its Application in Reprojection

Kuntal Sengupta, Ivan Poupyrev, Tatsumi Sakaguchi and Jun Ohya
ATR Media Integration & Communications Research Laboratories

Seitz and Dyer [8] have shown that given two images, the set of all views on the line joining the two optical centers can be generated by morphing. The view morphing idea therefore does not extend for arbitrarily placed virtual cameras. Shashua’s [7] reprojection algorithm uses trilinear forms of three perspective views for reprojection, the key assumption there is that the internal calibration parameters of the cameras are same.

Our approach is similar to [7] in the sense that it is neither based on 3D reconstruction, nor on epipolar line intersection. We build the theory for a new camera projection model, which is intermediate between the affine model and the pin hole model. The reprojection theory however is not limited to the “same internal calibration parameters” assumption. The paper is organized as follows. In Section II, we build the theory of affine coordinates based reprojection over three different scenarios. In Section III, we discuss experimental results in image based scene generation. In Section IV, we present our conclusions.

2 Properties of Affine Coordinates

In this section, we develop the theory of affine coordinate based reprojection over three different scenarios:

1. *Planar* : Here, all points in the scene belong to a plane in 3D. The perspective transform is approximated as an affine transform.
2. *Para Perspective Camera Model*: Points in the scene can lie anywhere, but we assume the camera projection model to be an orthographic projection into a plane followed by an affine trans-

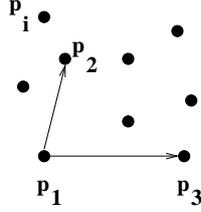


Figure 1: The points on the 3D plane, and the basis (p_1, p_2, p_3) .

form of these (projected) points as in the planar case.

3. *Our Camera Model:* Points lie anywhere, and we use the pin hole geometry of the cameras to project the points into a plane followed by an affine transform of these (projected) points as in the planar case. This camera model is intermediate between the Para Perspective model and the Full Perspective model.

Note that the *Planar* case is considered only for the purpose of successively building the reprojection theory from the para perspective model to our camera model.

2.1 Planar Case

Here, we use the theory of affine invariance originally proposed by Lamdan and Wolfson [5] for object recognition applications. Let $\{p_1, p_2, \dots, p_n\}$ be the projection of a set of points, all belonging to a 3D plane. Choosing (p_1, p_2, p_3) as the basis as shown in Fig. 1, the affine coordinates (α_i, β_i) of the i th point is invariant to affine transforms, where $p_i = p_1 + \alpha_i(p_2 - p_1) + \beta_i(p_3 - p_1)$. For 3D planar points, the transformation under perspective projection can be approximated as an affine transform [10]. Therefore the affine coordinates of any point on the plane are invariant over all possible sets of images.

2.2 Para Perspective Case

In [3], Jacobs simplifies the camera projection model as an orthographic projection into a plane followed by an affine transform of these (projected) points. For a set of points $(P_1, P_2, P_3, \dots, P_n)$, following [3], we construct a hypothetical plane passing through points P_1, P_2 and P_3 . We call it the basis plane, as shown in Fig. 2. The point P_4 is projected perpendicularly into the basis plane, and we call this projected point as p'_4 . The affine coordinates of p'_4 with respect to the basis (P_1, P_2, P_3) are (a_4, b_4) . Similarly, for the i th point, its projection on the basis plane is p'_i (with affine coordinates (a_i, b_i)). Let d_4 and d_i be the distances of point P_4 and p_i from the basis plane, respectively.

As in [3], for affine coordinates (α_4, β_4) , it can be shown that there is a viewpoint in which the projec-

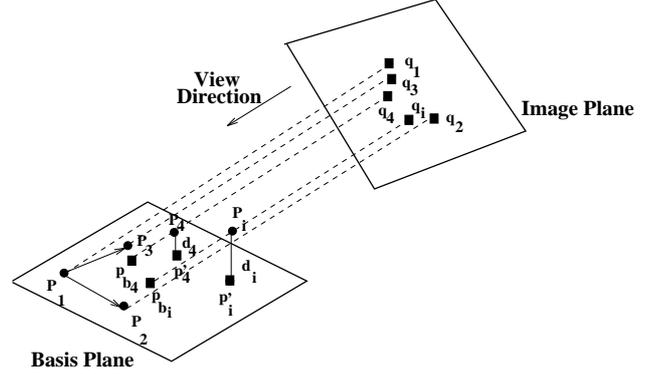


Figure 2: Projection of points into an image plane when the camera projection model is assumed to be an orthographic projection followed by an affine transform.

tion of the point P_4 has those affine coordinates. Let p_{b_4} be a point on the basis plane with affine coordinates (α_4, β_4) , for the basis (P_1, P_2, P_3) . The line passing through p_{b_4} and P_4 sets this viewing direction. This line meets the image plane (whose normal is parallel to the line) at a point q_4 . That is, q_4 is the image of P_4 . In a similar manner, we project P_1, P_2, P_3 into q_1, q_2 and q_3 , respectively, on this image plane. With (q_1, q_2, q_3) as the basis, one can easily observe that q_4 has the affine coordinates (α_4, β_4) , even when we subject the points on the image plane to an affine transformation (which includes translation, rotation, and scaling, to name a few).

Next, the affine coordinates (α_i, β_i) of the projections of the remaining points (for this given view direction) are computed as functions of (α_4, β_4) . Let p_{b_i} be the intersection point of the basis plane and the ray parallel to the viewing direction and passing through P_i . Let q_i be its projection on the image plane. As before, both p_{b_i} and q_i have the affine coordinates (α_i, β_i) when the basis chosen are (P_1, P_2, P_3) and (q_1, q_2, q_3) , respectively. Using similar triangles $P_4 p_{b_4} p'_4$ and $P_i p_{b_i} p'_i$, we have

$$(p_{b_i} - p_i) = \frac{d_i}{d_4} (p_{b_4} - p_4)$$

In terms of the affine coordinates, we express the equation as:

$$(\alpha_i, \beta_i) - (a_i, b_i) = \frac{d_i}{d_4} ((\alpha_4, \beta_4) - (a_4, b_4))$$

Considering the α coordinates only, $\alpha_i - a_i = \frac{d_i}{d_4} (\alpha_4 - a_4)$. Note that a_i, a_4, d_i and d_4 are constant over all possible images that can be generated for the given set of 3D points. Thus, for every possible image generated for (P_1, \dots, P_n) , the plot of (α_4, α_i) is a straight line with a slope $\frac{d_i}{d_4}$, as shown in Fig. 3. The slope of the line is indicative of how far P_i is from the basis plane. Similarly, plotting (β_4, β_i) leads to a line, with the exact same slope as in the case of α .

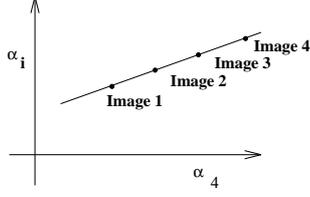


Figure 3: The plot of (α_4, α_i) over all possible images leads to a straight line, as shown here.

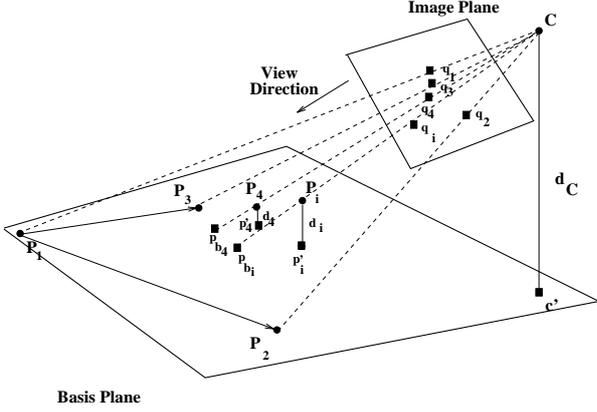


Figure 4: Projection of points into an image plane when the pin hole camera model is used for projection into a plane followed by an affine transform.

2.3 Our Camera Model

Here, we assume a regular pin hole camera geometry to project the points in 3D into a plane. This is followed by an affine transform of these (projected) points. Let $(P_1, P_2, P_3, \dots, P_n)$ be the set of 3D points not necessarily lying on a plane. We construct a hypothetical plane passing through points P_1, P_2 and P_3 as shown in Fig. 4 and go through a series of geometric constructions, as in the para perspective case.

Let q_i be the projection of P_i on the image plane. As before, both p_{b_i} and q_i has the affine coordinates (α_i, β_i) when the basis are (P_1, P_2, P_3) and (q_1, q_2, q_3) , respectively. Also, let c' and d_C be the projection of the camera center C and its distance from the image plane. Using similar triangle pairs $Cp_{b_i}c'$ and $P_i p_{b_i} p'_i$, and $Cp_{b_4}c'$ and $P_4 p_{b_4} p'_4$, we have:

$$\begin{cases} p_{b_i} - p'_i &= \frac{d_i}{d_C} (p_{b_i} - c') \\ p_{b_4} - p'_4 &= \frac{d_4}{d_C} (p_{b_4} - c') \end{cases} \quad (1)$$

Dividing the above two equations, and writing in terms of the affine coordinates, we have:

$$\frac{(\alpha_i, \beta_i) - (a_i, b_i)}{(\alpha_4, \beta_4) - (a_4, b_4)} = \left(\frac{d_i}{d_4} \right) \left(\frac{(\alpha_i, \beta_i) - (a_c, b_c)}{(\alpha_4, \beta_4) - (a_c, b_c)} \right)$$

In terms of the α coordinates only, we simplify it

to:

$$(\alpha_i - a_i) = \frac{d_i}{d_4} \left(\frac{\alpha_i - a_c}{\alpha_4 - a_c} \right) (\alpha_4 - a_4) \quad (2)$$

Let $\frac{d_i}{d_4} = k_i$. Writing the simultaneous equations for $i = 5, j$, and eliminating a_c , we have:

$$(\alpha_j - \alpha_4) (a_5 k_5 - a_4 k_5 k_j) + (\alpha_5 - \alpha_4) (-a_j k_j + a_4 k_5 k_j) + \alpha_j (\alpha_4 - \alpha_5) (k_5 k_j - k_j) + \alpha_5 (\alpha_4 - \alpha_j) (-k_5 k_j + k_5) = 0$$

Similarly, writing the equations for the β coordinate, we have:

$$(\beta_j - \beta_4) (b_5 k_5 - b_4 k_5 k_j) + (\beta_5 - \beta_4) (-b_j k_j + b_4 k_5 k_j) + \beta_j (\beta_4 - \beta_5) (k_5 k_j - k_j) + \beta_5 (\beta_4 - \beta_j) (-k_5 k_j + k_5) = 0$$

Let $C_1^j = (a_5 k_5 - a_4 k_5 k_j)$, $C_2^j = (-a_j k_j + a_4 k_5 k_j)$, $C_3^j = (b_5 k_5 - b_4 k_5 k_j)$, $C_4^j = (-b_j k_j + b_4 k_5 k_j)$, $C_5^j = (k_5 k_j - k_j)$, and $C_6^j = (-k_5 k_j + k_5)$. Note that C_1^j, \dots, C_6^j are invariants over all possible images, as they are functions of the scene structure. The above equations can also be written as:

$$\begin{bmatrix} (\alpha_j - \alpha_4) & (\alpha_5 - \alpha_4) & 0 & 0 & \alpha_j (\alpha_4 - \alpha_5) & \alpha_5 (\alpha_4 - \alpha_j) \\ 0 & 0 & (\beta_j - \beta_4) & (\beta_5 - \beta_4) & \beta_j (\beta_4 - \beta_5) & \beta_5 (\beta_4 - \beta_j) \end{bmatrix} [C_1^j \ C_2^j \ C_3^j \ C_4^j \ C_5^j \ C_6^j]^T = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \quad (3)$$

where the X^T indicates the transpose of a matrix X .

Thus, if we can locate and track 6 image points over n images, we get a total of $2n$ equations. We write the system of equations as

$$MC^{jT} = 0$$

where M is a $2n \times 6$ matrix, and $C^j = [C_1^j \dots C_6^j]$ spans the null space of M . To estimate C^j , we theoretically require only $2\frac{1}{2}$ images*. For a higher value of n , C^j is estimated as follows. Let M' be the $2n \times 5$ matrix constructed from M , by deleting the 6th column. Let m be a $2n \times 1$ matrix, whose entries correspond to those of the 6th column of M . Also, let $C_6^j = 1$, and $C^{jT} = [C_1^j \dots C_5^j]$. Using least squares,

$$C^{jT} = -(M'^T M')^{-1} M'^T m \quad (4)$$

In a novel image (we will label it as the $(n+1)$ th image) where we have located the images of the five reference points P_1, \dots, P_5 , we estimate the affine coordinates of the j th point using Eq. 3:

$$\hat{\alpha}_j^{n+1} = \frac{\alpha_4^{n+1} C_1 - (\alpha_5^{n+1} - \alpha_4^{n+1}) C_2 - \alpha_5^{n+1} \alpha_4^{n+1}}{C_1 - \alpha_5^{n+1} + (\alpha_4^{n+1} - \alpha_5^{n+1}) C_5} \quad (5)$$

*As number of images is always an integer, it means we need 3 images in total.

and

$$\hat{\beta}_j^{n+1} = \frac{\beta_4^{n+1} C_3 - (\beta_5^{n+1} - \beta_4^{n+1}) C_5 - \beta_5^{n+1} \beta_4^{n+1}}{C_3 - \beta_5^{n+1} + (\beta_4^{n+1} - \beta_5^{n+1}) C_5} \quad (6)$$

where $(\alpha_i^{n+1}, \beta_i^{n+1})$ are the affine coordinates of the projection of P_i in the $(n+1)$ th image.

Therefore, if we know the location of 5 reference points in $(n+1)$ images, and if we have located an arbitrary point in n of these images, we can uniquely estimate its position in the $(n+1)$ th image using our camera model. Note that the minimum (integral) value of n is 3. One may argue that since we know the location of the 6 points (5 reference points and the j th point) in three images, we can estimate the perspective structure of each of the six points using the technique in [6], and use this information to reproject the j th point into the $(n+1)$ th image. However, one may note that there are three sets of solutions in the method described in [6], whereas we obtain an *unique* solution, albeit with an approximate camera model.

We observe that $[C_1^j \dots C_6^j]$ contains rich information about the scene structure. For example, with $C_6^j = 1$, $C_5^j = \frac{k_5 k_i - k_j}{-k_5 k_j + k_5}$. The parameter C_5^j decreases monotonically with k_j , the later being the relative depth of the point P_j (with respect to P_4) from the basis plane. This observation is used in novel view generation for detection of occluded points.

A Special Case

We rewrite Eq. 2 as

$$(\alpha_i - a_i) = \frac{d_i}{d_4} \alpha'_4 \quad (7)$$

where

$$\alpha'_4 = \left(\frac{\alpha_i - a_c}{\alpha_4 - a_c} \right) (\alpha_4 - a_4) \quad (8)$$

Thus, the plot of (α'_4, α_i) leads to a straight line. We observe a similar property for the β coordinate.

Let the geometry of the five reference points in the 3D space (P_1, \dots, P_5) be such that P_1, P_4 and P_5 are collinear and $P_1 P_4$ (or $P_1 P_5$) is perpendicular to the plane containing P_1, P_2 and P_3 . Also, let $|P_5 P_1| = k |P_4 P_1|$. Thus $a_4 = a_5 = 0$. In an image, using Eqs. 7 and 8, we have:

$$a_c = \frac{\alpha_4 (1 - k)}{\left(1 - k \frac{\alpha_4}{\alpha_5} \right)} \quad (9)$$

Using Eqs. 7 and 9, and observation from *only* two images, we draw the straight line in the α space. Let the equation of the line be $\alpha_i = \kappa_0 \alpha'_4 + \kappa_1$. Using Eqs. 7 and 8, we have

$$\hat{\alpha}_i^{n+1} = \frac{-\frac{\kappa_0 \alpha_c^{n+1} \alpha_4^{n+1}}{\alpha_4^{n+1} - a_c^{n+1}} + \kappa_1}{1 - \frac{\kappa_0 \alpha_4^{n+1}}{\alpha_4^{n+1} - a_c^{n+1}}} \quad (10)$$



Figure 5: (a) One of the images from the input sequence (b) The target image used to verify the quality of reprojection. A few points, marked in green color are hand picked and reprojected onto the target image.

where $(\hat{\alpha}_i^{n+1}, \hat{\beta}_i^{n+1})$ are the coordinates of the projection of P_i in the novel image (also labeled as the $(n+1)$ image). Thus, reprojection is achieved with only two input images, and location of the five *special* reference points across the two input image, and the novel image.

Novel View Generation

The details of the novel view generation algorithm is outside the scope of the paper, however we discuss it here briefly. It first assumes that dense correspondences have been computed between the n image sequences. For stereo images, we use a correlation based stereo matching for this purpose. For monocular image sequences, we use an optical flow algorithm. Given the location of the reference points in the input images and the novel image, the remaining points are reprojected using the algorithm described in the paper. We use a polygonal representation of the image to do rendering [9].

3 Experimental Results

Using a Sony Digital HandyCam, we took a monocular image sequence of an office scene, the first frame of the sequence is shown in Fig. 5(a) (For a few other images, the reader can look at Fig. 6(b)). For the purpose of verification of the estimation accuracy, we chose a real image (which does not belong to the image sequence) which we will call as the *target image*. We arbitrarily hand picked five reference points across the input image sequence and the target image. We then reproject a few other points (numbered and illustrated in green color on the images) into the target image. Looking at Fig. 6(b), we observe that the estimated location of the reprojected points is comparable to their actual location in the target image. Small estimation errors creep in from the inaccuracies in tracking over the input image sequences, and also due to the assumption in the camera projection model.

Using the reprojection and novel view generation



Figure 6: (a) A few images from the input sequence. (b) Images from the virtual walkthrough sequence generated using our algorithm.

approach discussed in the earlier section, we generate a new synthetic image sequence, which corresponds to the “virtual walkthrough”, the virtual camera moving closer to the person in the scene. The input image sequence is illustrated in Fig 6(a), and the synthetic image sequence is in Fig. 6(b). Note that the quality of the synthetic images are quite good, although straight edges come out bent at certain places. Such inaccuracies can be attributed to errors in tracking and due to polygonization in the rendering [†].

For an example where we have stereo images as inputs as shown in Fig. 7 (a), some of the novel views generated are shown in Fig. 7(b).

4 Conclusions

In this paper, we present an affine coordinate based reprojection scheme for a new camera model, which is intermediate between the affine and the pin hole model. It does not take into account assumptions such as “same internal calibration” parameters, and we need to track only five reference points over the images. We also show how our method can be used for image based rendering in different domains, such as (a) Virtual walkthrough application, where the input is a monocular image sequence, (b) Novel view generation where the input is a stereo pair.

Also, we note that we can estimate the structure of a scene, as the parameter set C^j (estimated using Eq. 4) contains rich information about the geometry of the points. Algorithms discussed in [6, 11] discusses structure estimation method for a more accurate pin hole camera model. However, the structure estimated there is not unique (there are three set of solutions). Our method, although slightly inaccurate, guarantees an unique solution.

[†]We did not discuss the details of the polygonization process due to lack of space.

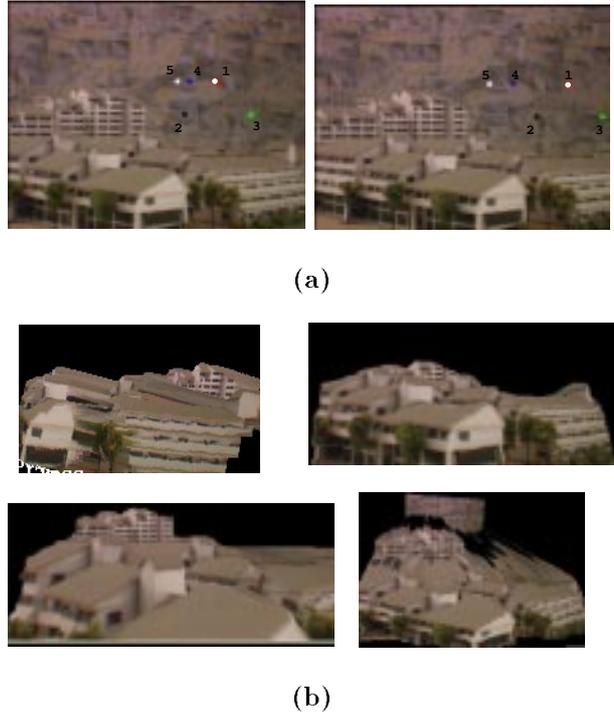


Figure 7: (a) The stereo pairs used for the experiment of novel view generation. (b) The novel views generated.

References

- [1] O. Faugeras, “What can be seen in three dimensions with an uncalibrated stereo rig?” *Proceedings of ECCV*, pp. 563–578, 1992.
- [2] O. Faugeras and L. Robert, “What Can Two Images Tell Us About a Third One?” *IJCV*, vol.18, pp. 6–19, 1996.
- [3] D. W. Jacobs, “Space Efficient 3D Model Indexing,” *Proceedings of CVPR*, pp. 439–444, Jun. 1992.
- [4] T. Kanade, A. Yoshida, K. Oda, H. Kano and M. Tanaka, “A stereo machine for video rate dense depth mapping and its new applications,” *Proceedings of CVPR*, pp. 196–202, Jun. 1996
- [5] Y. Lamdan and H. Wolfson, “Geometric Hashing: A General and Efficient Model-Based Recognition Scheme,” *IEEE Proc. of Robotics and Automation*, pp. 238–249, 1988.
- [6] L. Quan, “Invariants of six points and projective reconstruction from three uncalibrated cameras,” *IEEE Trans. PAMI*, pp. 34–46, vol. 17, no. 1, 1995.
- [7] A. Shashua and S. Avidan, “The Rank 4 Constraints in Multiple (≥ 3) View Geometry,” *Proc. ECCV*, pp. 196–206, Apr. 1996.

- [8] S. Seitz and C. Dyer, "Toward Image-Based Scene Representation Using View Morphing," *Proceedings of the 13th ICPR*, pp. 84–89, 1996.
- [9] K. Sengupta and J. Ohya, "Novel Scene Generation, Merging and Stitching Views Using The 2D Affine Space", Proc. IEEE Int. Conference on Multimedia Computing and Systems, Ottawa, pp. 602–603 , Jun. 1997.
- [10] R. Haralick and L. Shapiro, *Computer and Robot Vision, Volume II*, Addison-Wesley Publishing Company, pp. 468–469, 1993.
- [11] Z. Zhang, Q. Luong and O. Faugeras, *Motion of an uncalibrated stereo rig: Calibration and Metric Reconstruction*, *IEEE Trans. PAMI*, pp. 103–113, vol. 12, no. 1, 1996.