6—1

Novel View Generation Using Affine Coordinates

Kuntal Sengupta and Jun Ohya
ATR Media Integration and Communications Research Laboratories
2-2 Hikaridai, Seika cho, Soraku gun, Kyoto 619-02, Japan
{kuntal,ohya}@mic.atr.co.jp

Abstract: In this paper we present an algorithm to generate new views of a scene, starting with images from weakly calibrated cameras. Errors in 3D scene reconstruction usually gets reflected in the quality of the new scene generated, so we seek a direct method for reprojection. In this paper, we use the knowledge of dense point matches and their affine coordinate values to estimate the corresponding affine coordinate values in the new scene. We borrow ideas from the object recognition literature, and extend them significantly to solve the problem of reprojection. Unlike the epipolar line intersection algorithms for reprojection which requires at least eight matched points across three images, we need only five matched points. The theory of reprojection is used with hardware based rendering to achieve fast rendering. We demonstrate our results of novel view generation from stereopairs for arbitrary locations of the virtual camera.

I. 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.

In this paper, we refer to the images from the real, "weakly" calibrated cameras as real images(for example, stereo pairs) , and those from the virtual camera as novel/virtual images. By weak calibration, we mean that we know the epipolar geometry of the stereo pair. For novel view generation, to start with, we assume that we have already solved the problem of dense point match correspondence among the real images. In this paper, we show how these matched points can be *reprojected* into a novel image.

There are approaches in the literature which solves the problem of scene reprojection either by extracting the 3D structure of the scene either using the strong camera calibration parameters, or by assuming that the objects undergo 3D affine or projective transformation [4, 5]. This is followed by the projection of the 3D points into the new image. As mentioned by Shashua in [1], 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 [6].

Most recently, Seitz and Dyer [3] 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. In that sense, Shashua's [1] reprojection algorithm using trilinear forms of three perspective views seem to be the most technically correct approach. However, we are not very sure about its time performance. Our approach is similar to [1] in the sense that it is neither based on 3D reconstruction, nor on epipolar line intersection. The affine coordinate based scene reprojection theory presented in this paper is much simpler, and is based on the work reported in [2]. Here, the author shows that the set of 2D images produced by a group of 3D point features of a rigid model can be optimally represented by two lines in a high dimensional $\alpha - \beta$ (affine) space.

The assumption of the of orthogonal projection does not produce encouraging results, so we replace it with a pin hole camera projection model. We study the properties of the affine coordinates of matched points for this model, and use them for novel view generation.

The paper is organized as follows. In Section II we discuss the affine coordinate based reprojection approach, and show how the theory can be used for novel view synthesis. We present the experimental results in Section III, followed by a section on conclusions.

II. Scene Reprojection and Novel View Generation

In this section, we sequentially develop the idea of affine coordinate based reprojection over two different scenarios:

1. *Planar Case:* Here, all points in the scene belong to a plane in 3D. The perspective transform is approximated as an affine transform.
2. *Pin Hole Case:* 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.

A. Planar Case

Here, we use the theory of affine invariance originally proposed by Lamdan and Wolfson [7] for object recognition applications. Let $\{p_1, p_2, \ldots, p_n\}$ 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 ith point is invariant to affine transforms, where

$$p_i = p_1 + \alpha_i (p_2 - p_1) + \beta_i (p_3 - p_1)$$

Approximating the transformation under perspective projection as an affine transform, we can say that the affine coordinates of any point on the plane are invariant over all possible sets of novel images.

Scene reprojection and novel view generation is quite simple when we assume that all points in the world belong to a plane. We need only one scene to reproduce any other scene. We choose an affine basis and a good engineering solution would be to show the real camera a structure like the one in Fig. 1(b)[1]. Here, P_1P_2 is at right angles to P_1P_3, and $P_1P_2 = P_1P_3$. Without loss of generality, let P_1, P_2 and P_3 project to points p_1, p_2 and p_3, respectively, in the real image. Also, given the 3×4 calibration matrix of the virtual camera, we project P_1, P_2 and P_3 into points (in its image plane) p_1^v, p_2^v and p_3^v, respectively. For this, we assume that the world coordinates of P_1, P_2 and P_3 are (0, 0, 0), (1, 0, 0), and(0, 1, 0), respectively. For a point p_i^v in the novel image, we compute its affine coordinates (α_i^v, β_i^v) with respect to the basis (p_1^v, p_2^v, p_3^v). The corresponding point in the real image is p_i, where:

$$p_i = p_1 + \alpha_i^v (p_2 - p_1) + \beta_i^v (p_3 - p_1) \quad (1)$$

The point p_i^v in the virtual image assumes the same greyscale/color values as the point p_i.

B. Pin Hole Case

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 $(P_1, P_2, P_3, \ldots, P_n)$ 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. 2. We call it the *basis plane*. Also, for the point P_4, we drop a perpendicular on the basis plane and construct point p_4', whose affine coordinates are (a_4, b_4) for the basis (P_1, P_2, P_3). We similarly construct the point p_i'.

For affine coordinates (α_4, β_4), it can be shown that there is a viewpoint in which the projection 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 the line on which the camera center C can lie on, which we pick arbitrarily. We also choose the viewing direction arbitrarily. The line $P_4p_{b_4}$ meets the image plane 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, we observe that q_4 has the affine coordinates (α_4, β_4), with the affine transform approximation of perspective transform.

Let q_i be the projection of P_i on the basis 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),

[1] Any three non collinear point in space will do the job, we simply make its geometry as simple as possible.

respectively. Also, let c' and d_C be the projection of the camera center C and its distance from the image plane.

Using similar triangles $Cp_{b_i}c'$ and $P_ip_{b_i}p'_i$, we have:

$$p_{b_i} - p'_i = \frac{d_i}{d_C}(p_{b_i} - c') \quad (2)$$

Using similar triangles $Cp_{b_4}c'$ and $P_4p_{b_4}p'_4$, we have.

$$p_{b_4} - p'_4 = \frac{d_4}{d_C}(p_{b_4} - c') \quad (3)$$

Thus, from Eqns. (2) and (3), we have:

$$\frac{p_{b_i} - p'_i}{p_{b_4} - p'_4} = \left(\frac{d_i}{d_4}\right)\left(\frac{p_{b_i} - c'}{p_{b_4} - c'}\right)$$

In terms of the affine coordinates, the above equation is written as:

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

In terms of the α coordinates only, we simplify it to:

$$(\alpha_i - a_i) = \frac{d_i}{d_4}\alpha'_4 \quad (4)$$

where

$$\alpha'_4 = \left(\frac{\alpha_i - a_c}{\alpha_4 - a_c}\right)(\alpha_4 - a_4) \quad (5)$$

Although a_i, a_4, d_i and d_4 are constant over all possible images that can be generated, a_c is dependent on the camera parameters. However, if we know a_c and a_4 *a-priori*, then we can easily compute α'_4 for a given image. Thus, for every possible image, the plot of (α'_4, α_i) is a straight line with a slope of $\frac{d_i}{d_4}$. This property is illustrated in Fig. 3. The line in the β space has the same slope as the line in the α space.

Also, note that the slopes of the line corresponding to the ith point is directly proportional to the distance of the point from the basis plane.

Unlike the planar case, we need to know the location of a 3D point in two or more images (for example, stereo pairs) to project it into a new image. Let the two given images be I_1 and I_2. For novel view generation, we assume the knowledge of dense point correspondence between these two images. For a point p_i^1 in image I_1, let the corresponding point in I_2 be p_i^2. We need four reference points (three points to create the basis and a fourth point) to generate the lines in α and β space. Without loss of generality, let the reference points be $p_1^j, p_2^j, p_3^j, p_4^j$ in the image I_j $(j = 1, 2.)$. To make things simpler, we choose these points as the images of the points P_1, P_2, P_3 and P_4. P_1, P_2 and P_3 are shown in Fig. 1(b) and the line P_1P_4 is perpendicular to the plane containing P_1, P_2 and P_3. Also, $|P_1P_4| = |P_1P_2| = |P_1P_3| = |P_2P_3| = 1$. We show this structure (simultaneously) to the two cameras before the experiment, and record the coordinate values of their projections $p_1^j, p_2^j, p_3^j, p_4^j$ $(j = 1, 2)$. Now, for points p_4^j and p_i^j in image I_j, let its affine coordinates be (α_4^j, β_4^j) and (α_i^j, β_i^j), respectively. The line in the (2 dimensional) α space discussed earlier passes through the points $(\alpha_4^{1\prime}, \alpha_i^1)$ and $(\alpha_4^{2\prime}, \alpha_i^2)$.

Note that for the jth image, to compute $\alpha_4^{j\prime}$ using Eqn. (5), we need to know a_c^j and a_4. For the chosen geometry of $P_1, \ldots, P_4$, a_4 is zero.. To compute a_c^j, we need a fifth control point (say P_5). For our convenience, we choose P_5 such that it is collinear with P_4 and P_0. Let $|P_5P_0| = k|P_4P_0|$, where $|P_5P_0|$ denotes the Euclidean distance between points P_5 and P_0, and so on. Let P_5 project to the point p_5^j in the jth camera. Let its affine coordinates be (α_5^j, β_5^j). Since $a_5 = 0$, using Eqns. (4) and (5), we have:

$$a_c^j = \frac{\alpha_4^j(1-k)}{\left(1 - k\frac{\alpha_4^j}{\alpha_5^j}\right)} \quad (6)$$

Without loss of generality (for generating the virtual image), we assume that the coordinate values of P_1, P_2, P_3, P_4, and P_5 are (0 0 0), (1 0 0), (0 1 0), (0 0 1), and(0 0 k), respectively. These points are projected to the novel image using the 3 × 4 perspective transformation matrix of the virtual camera. Let these points be $(p_1^v, \ldots, p_5^v)$. Reprojecting the ith point p_i^1 is accomplished by computing the affine coordinates (α_i^v, β_i^v), with (p_1^v, p_2^v, p_3^v) as the basis. Let $\alpha_i = \kappa_0\alpha'_4 + \kappa_1$ be the equation of the line in the α space for the ith point. Using Eqns. (4) and (5), we have

$$\alpha_i^v = \frac{-\frac{\kappa_0 a_c^v \alpha_4^v}{\alpha_4^v - a_c^v} + \kappa_1}{1 - \frac{\kappa_0\alpha_4^v}{\alpha_4^v - a_c^v}} \quad (7)$$

where α_4^v and a_c^v are the α component of the affine coordinates of p_4^v and the projection of the virtual camera center. We compute β_i^v similarly. All points in I_1 are reprojected to the virtual image using this technique.

Note that two or more points in I_1 can map to a point in the virtual image. To resolve this, we can use the slope of the line information to decide which point

to choose. A point with a larger slope value is further away from the basis plane, and hence closer to the virtual image, assuming that the virtual camera is *looking into* the basis plane. Also, there is the issue of *filling up the gaps* in the novel image, since our process does not guarantee that every point in it will be mapped to by a point in I_1. We address both these problems using standard ideas in computer graphics, like polygonization. We essentially divide I_1 into squares, each with sides of one pixel length as shown in Fig. 4. For the point p_i^1 shown in the figure, the x and y coordinates are the x and y coordinates of p_i^v, respectively. The z coordinate value is the negative of the slope of the line in the affine space, and the texture value is identical to that of p_i in I_1. This process is repeated for the remaining three points in the square, and for all squares in the grid. Next, we use the graphics hardware to render this polyhedra under an orthographic projection without any translation, rotation or scaling. This retains the x and y coordinates of the vertices in the polyhedra. This process of rendering should not be confused with the 3D reconstruction of the scene and rendering, because the polyhedral representation of I_1 is not the $3D$ structure of the scene. It is completely derived from the reprojected coordinate values.

III. Experimental Results

To verify the theory of scene reprojection using affine coordinates, we experimented on three real images of a scene simply taken by moving the IndyCam mounted on an SGI workstation. The images are shown in Fig. 5. Although in our previous discussions, the third image has always been the virtual image, it is not difficult to see that the theory works when the third one is a real image, provided we have five matched points (which are essentially projections of $P_1, \ldots, P_5$) over the images. From the checkered block, we chose five matched points over the three images, marked by 'x'. We also hand picked seven match points between the first two images as shown by black dots in Fig. 5(a) and (b). Fig.5(c) illustrates how these matched points project into the third image while we use the pin hole model of the camera. We observed that our theory of reprojection provides a marked improvement over the reprojection results under Jacob's camera model in [2].

The images shown in Fig. 6(a) are two images from a multiple baseline stereo configuration. The images were already rectified, so we implemented a straight forward correlation based stereo matching algorithm [8] to generate the disparity map and the dense point match information. The five reference points in the two images are shown as dots. We use the theory in Sec. II.B. to generate new views of the scene, as shown in Fig. 6(b). At present, the novel view generation algorithm runs almost real time on an SGI Onyx workstation. We have compared our results with the standard 3D reconstruction and rendering based novel view generation. With the exception of a scale factor, the results from these two approaches are not significantly different.

IV. Conclusions

In this paper, we present an algorithm for scene reprojection and novel view generation using properties of affine coordinates in sets of images that can be produced by a collection of 3D points. This method requires only five matched points across three images, of which two are real images and the third one is the novel image. We believe that our theory can be extended to other applications such as generating epipolar lines, view stitching and merging real and virtual objects. Note that the computations (Eq. (7)) for scene reprojection is quite simple. Also using the graphics hardware leads to almost real time performance of the novel view generation process on still scenes. For dynamic scenes, the biggest bottleneck is in obtaining dense point matches at frame rate. With standard stereo matching algorithms available in the literature, this can only be accomplished by designing a special purpose hardware.

References

[1] A. Shashua, "Algebraic Functions for recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 17, pp. 779–789, Aug. 1995.

[2] D. W. Jacobs, "Space Efficient 3D Model Indexing," *Proceedings of CVPR*, pp. 439–444, Jun. 1992.

[3] S. Seitz and C. Dyer, "Toward Image-Based Scene Representation Using View Morphing," *Proceedings of the 13th ICPR*, pp. 84–89, 1996.

[4] O. Faugeras, "What can be seen in three dimensions with an uncalibrated stereo rig?" *Proceedings of ECCV*, pp. 563–578, 1992.

[5] K. Kutulakos and J. Vallino, "Affine Object Representation for Calibration-Free Augmented Reality," *Proc. VRAIS*, pp. 25–36, 1996.

[6] O. Faugeras and L. Robert, "What Can Two Images Tell Us About a Third One?" *IJCV*, vol.18, pp. 6–19, 1996.

[7] Y. Lamdan and H. Wolfson, "Space Efficient 3D Model Indexing," *IEEE Proc. of Robotics and Automation*, pp. 238–249, 1988.

[8] M. Okutomi and T. Kanade, "A Multiple baseline stereo,", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 15, pp. 353–363. Aug. 1993.

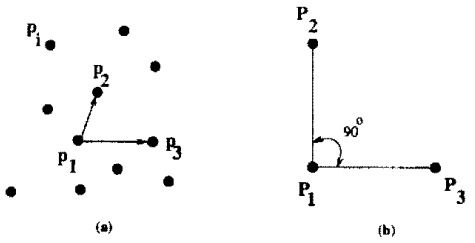

Figure 1: (a) The points on the 3D plane, and the basis (p_1, p_2, p_3). (b) The 3D reference points (P_1, P_2, P_3) which project to form the basis (p_1, p_2, p_3).

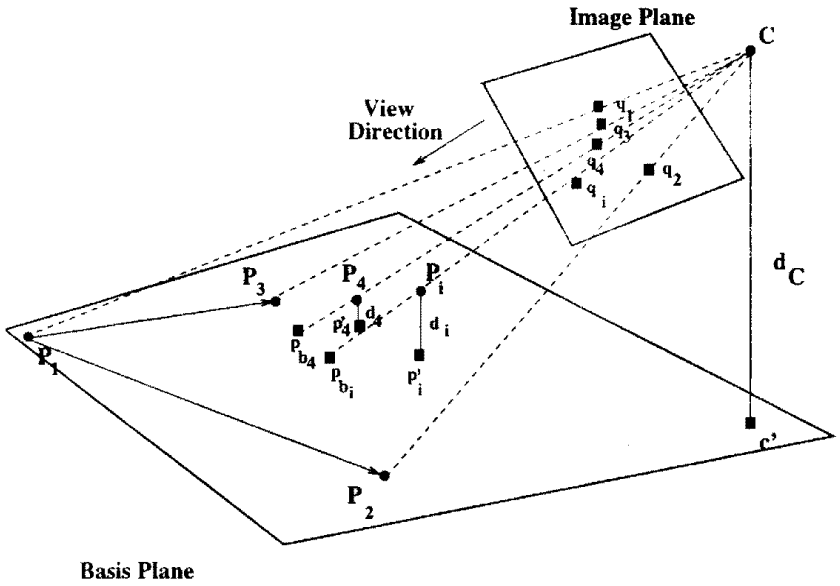

Figure 2: Projection of points into an image plane when the pin hole camera model is used for projection into a plane followed by an affine transform.

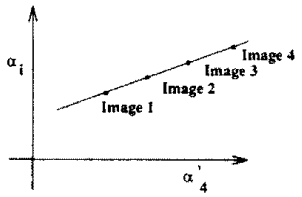

Figure 3: The plot of (α'_4, α_i) over all possible images leads to a straight line, as shown here.

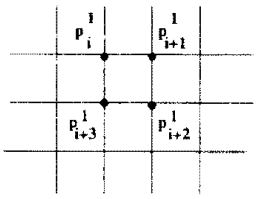

Figure 4: The construction of the grid for the polygonization process in generating novel views.

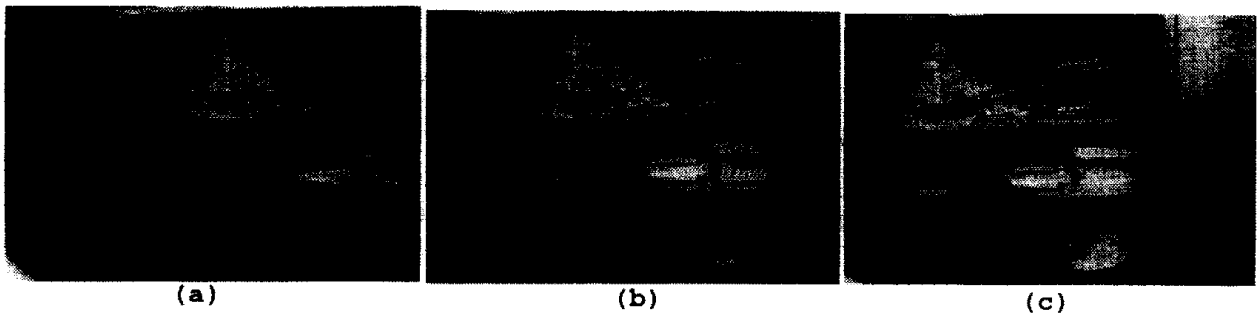

Figure 5: Results of the reprojection algorithm. The three images are shown in **(a)**, **(b)** and **(c)**. The 'x' represents the reference points. The dots in **(a)** and **(b)** represent the points to be reprojected. The dots in **(c)** represent the location of the reprojected points using our theory.

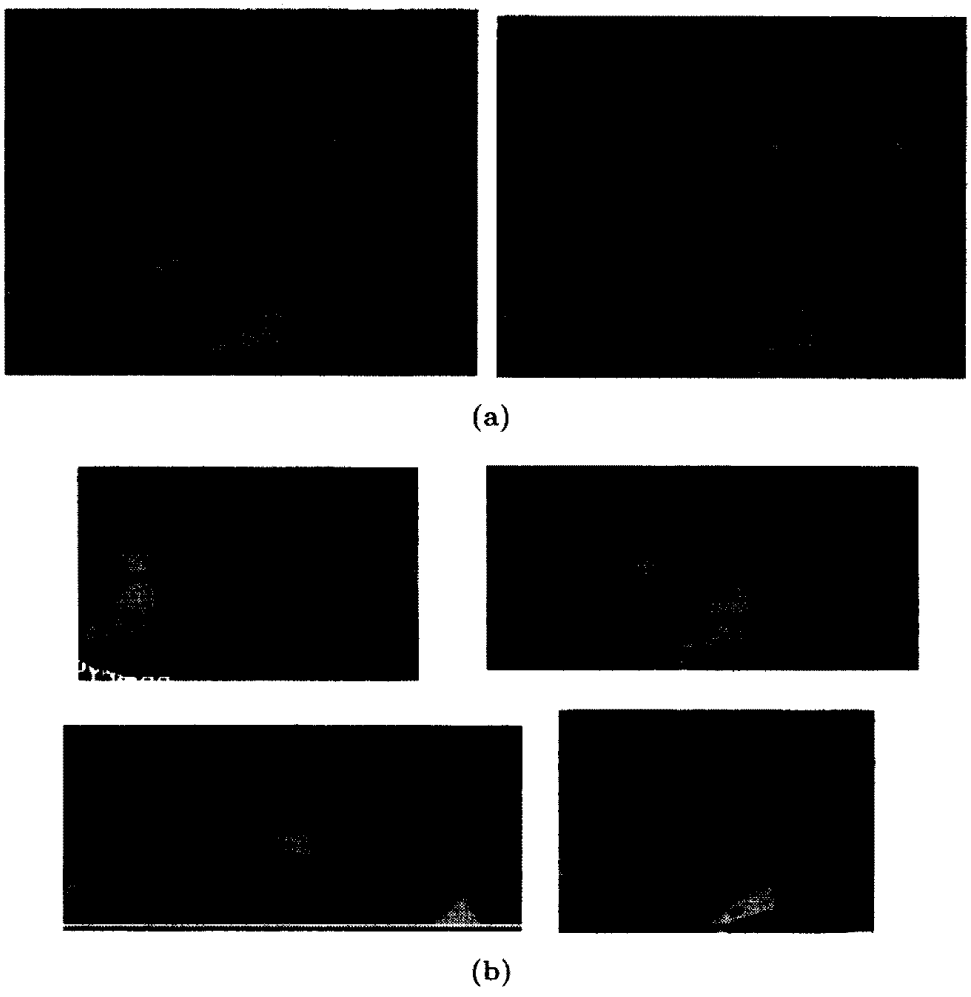

Figure 6: (a) The stereo pairs used for the experiment of novel view generation. (b) The novel views generated.