Novel Scene Generation, Merging and Stitching Views Using the 2D Affine Space

Kuntal Sengupta and Jun Ohyaa
ATR Media Integration and Communications Research Laboratories
2-2 Hikaridai, Seika cho, Soraku gun, Kyoto 619-02, Japan

Abstract
In this paper we present an algorithm to generate new views of a scene, starting from stereo images. Errors in 3D scene reconstruction usually get reflected in the quality of the new scene generated, so we seek a direct method for reprojen. 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.

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 abstract, we present a direct approach to scene reprojection, contrary to the traditional 3D reconstruction and rendering strategy [2]. It is based on [1], where the author shows that the set of 2D images produced by a group of 3D point features can be optimally represented by two lines in the high dimensional affine space.

II. Affine Coordinate Based Reprojection and Applications
We assume a regular pin hole camera geometry, instead of the para perspective model in [1], 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 \(P_1, P_2, \text{and } P_3\) 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 \(P_4'\) and \(c'\), their affine coordinates being \((a_4, b_4)\) and \((a_c, b_c)\), respectively. Let \(q_1, q_2, \ldots, q_n\) be the images of the 3D points in the image. Using \((q_1, q_2, q_3)\) as the basis, let the affine coordinates of \(q_j\) be \((\alpha_j, \beta_j)\), for \(j = 1, \ldots, n\).

Using standard geometric theorems in similar triangles, we can show that:

\[
\frac{\alpha_i - a_i}{d_i} = \frac{\alpha_j' - a_i}{d_j}
\]

where \(\alpha_i = \frac{a_i - a_4}{a_i - a_4}\) and \(d_j\) is the distance of \(P_j\) from the basis plane. Thus the plot of \((\alpha_j', \alpha_i)\) over all possible images fall on one straight line. The expression for the \(\beta\) coordinate can be written similarly. The straight line property can be used along with the dense point match information between stereo pairs to generate novel views.

Let the two input 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 \(P_1'\) in image \(I_1\), let the corresponding point in \(I_2\) be \(P_2'\). We need four reference points (three points to create the basis and a fourth point) to generate the lines in \(\alpha\) and \(\beta\) space. Let the reference points be \(P_1', P_2', P_3', P_4'\) 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 \(P_4\), where \(P_1P_2\) is perpendicular to \(P_1P_3\), and \(P_4P_1\) is perpendicular to the plane containing \(P_1, P_2, P_3\). Also, let \(|P_1P_4| = |P_1P_2| = |P_3P_1| = |P_2P_3| = 1\). We show this structure (simultaneously) to the two cameras before the experiment, and record the coordinate values of their projections \(P_1', P_2', P_3', P_4'\) \((j = 1, 2)\). Now, for points \(P_1'\) and \(P_2'\) in image \(I_j\), let its affine coordinates be \((\alpha'_j, \beta'_j)\) and \((\alpha_{j'}, \beta_{j'})\), respectively. The line in the (2 dimensional) \(\alpha\) space discussed earlier passes through the points \((\alpha_{1'}, \alpha_1')\) and \((\alpha_2, \alpha_1)\). To obtain \(\alpha_j'\) and \(\beta_j'\), we need to compute \(a_c\) and \(b_c\). For this, we use a fifth control point \(P_5\) collinear with \(P_1\) and \(P_4\), and \(|P_5P_6| = k|P_4P_6|\). We record its position in \(I_1\) and \(I_2\), respectively.

Without loss of generality (for novel view synthesis), 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 $3 \times 4$ perspective transformation matrix of the virtual camera. Using these five reference points and the straight line property of the affine coordinates shown in Eq. (1), we reproject the matched points in the two images to the novel image. We fill up the gaps in the novel image using the graphics hardware.

For merging real and synthetic objects, the polygonal (wavefront) representation is first defined with respect to the axis system of the reference points, followed by the simultaneous rendering of the reprojected points and the synthetic object. Two views are stitched by reprojecting one of the images into the other image.

### III. Experimental Results

The images shown in Fig. 2(a) are two images from a multiple baseline stereo configuration. We use a correlation based stereo matching algorithm to obtain the dense point match information. Example of novel views generated are shown in Fig. 2(b). An example of merging real and virtual objects is shown in Fig. 3. The result of stitching the two scenes in Fig. 2(a) appears in Fig. 4.

**References**


---

**Figure 1:** 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 2:** (a) The stereo pairs used for the experiment of novel view generation. (b) The novel views generated.

**Figure 3:** An example of merging the above image with a face model.

**Figure 4:** An example of stitching the stereo pairs.