Novel Views from Non-calibrated Stereo

M. A. Akhloufi, P. Cohen and V. Polotski
Groupe de Recherche en Perception et Robotique
Ecole Polytechnique de Montreal
P.O. Box 6079, Station Centre-Ville,
Montreal (Qc), Canada, H3C 3A7.
(akhloufi, cohen, polotski)@ai.polymtl.ca

Abstract

In this paper we present a new approach to the synthesis of novel views from two images given by an uncalibrated stereo system. Unlike methods based on inferring the 3D structure of the scene or on using dense correspondence between source images to produce a new synthesized view, we use epipolar constraints associated with two cameras configuration and represented by a ***fundamental matrix*** *to reproject corresponding features in the image plane of the view to be synthesized. This requires only sparse correspondence between features in the source images.* ***Perspective image warping*** *is used to render the remaining dense set of image points via texture mapping. This new approach allows interactive view synthesis in applications such as: immersive telepresence systems, virtual and augmented reality and telerobotics. Only an initialization process which consists in matching features between the source views is needed. The efficiency of the method is illustrated on images of synthetic and real scenes.*

1 Introduction

In recent years, researchers from the traditionally separate fields of computer vision and computer graphics have been working on a common problem, namely the development of tools that permit a realistic rendering from a sparse set of images.

In classical approaches novel views are rendered from appropriate reconstructed 3D models. In computer graphics the object is first represented using a 3D modeler. Then a texture is mapped to add realism to the scene [22, 8]. In computer vision the classical approach is similar. First, a 3D model is reconstructed from a set of input images. Then, a texture extracted from the images is mapped onto the model [6]. Novel views are obtained from a final model at a given viewpoint. Many techniques have been proposed for 3D model reconstruction [13, 12, 6]. However, it is known that 3D model reconstruction is complex, time consuming and prone to errors [11]. The second step of the procedure, the rendering from the 3D models, may also involve intensive computation for achieving a visually realistic image.

To overcome these problems, an emerging field of research, *image based rendering*[17, 20] has been recently introduced in computer vision and computer graphics. It is currently gaining relevent interest. One of the first works in image-based rendering was proposed by Chen and Williams [5], who developed the *QuickTime VR* system. It consists mainly of mosaicking a set of images taken from a camera rotating about the axis passing through its principal point. These views are then stitched together prior to be reprojected, via cylindrical mapping, on a common cylindrical reference frame. The user can interactively move within the cylinder to display different views in the panoramic sequence captured by the rotating camera. The system suffers from the restrictions associated with the camera motion used to capture the views. Other techniques use view interpolation between source images to synthesize the novel view [4], but do not achieve a perspective rendering and need a dense set of correspondences between input images.

Introducing geometrical constraints leads to methods which produce geometrically valid pixel reprojections and, hence, synthesize views that are close to the desired real ones. In this category, we may distinguish three techniques: *(1)* Faugeras and *al.* [19, 7, 16] use the epipolar constraint captured by the *fundamental matrix* to reproject the corresponding points in the new image plane from dense correspondence between the source images. *(2)* Avidan and Shashua [3] use *trilinear tensors* to synthesize a novel view, also from dense correspondence. *(3)* In [21] the author propose

the use of projective invariants to transfer the points in the third view. Seitz and Dyer [20], also use the fundamental matrix in a different manner that the one in [19]. The authors in [20] use linear morphing between rectified views to synthesize a virtual rectified view which is, then, reprojected to the desired image plane. Dense correspondence is needed to morph each pair of corresponding pixels between two source views via a linear cross-dissolve function.

In this paper, we present an approach which has several advantages over classical techniques. Here no explicit 3D model is needed and given two images of the scene, only a sparse correspondence between key points is used to synthesize the third view. Also, to insure a geometrically valid representation we use the epipolar constraints associated with the fundamental matrix to reproject the corresponding points from the two source images into their coorresponding points in the novel view. Finally, perspective image warping render the remaining image points in order to achieve a photo-realistic view.

The paper is organized as follows. In section 2, we introduce basic concepts of epipolar geometry, and an overview of the principal methods used to compute the fundamental matrix. Then, we present a new linear method for fundamental matrix computation. Section 3 shows how to obtain the features in the novel view from existing ones via reprojection. In section 4, we describe the perspective texture mapping technique used in the rendering process. Section 5 and 6 present the experimental results and few concluding remarks, respectively.

2 Fundamental matrix

Given two perspective images of a scene taken by two pinhole camera, epipolar geometry captures the basic relation between these images, described by a 3×3 matrix called the *fundamental matrix*. This matrix incorporates all the geometrical information of the two cameras [7, 18]. In this section we review the basic concepts of epipolar geometry relevant to the view generation problem.

2.1 Theory

In stereo vision, the right camera position and orientation with respect to the left camera are usually represented by a 4×4 homogeneous matrix (Figure 1)

$$\begin{bmatrix} \mathbf{R} & \mathbf{t} \\ 0 & 1 \end{bmatrix}$$

where $\mathbf{R}$ is a rotation matrix and $\mathbf{t}$ is a translation vector.

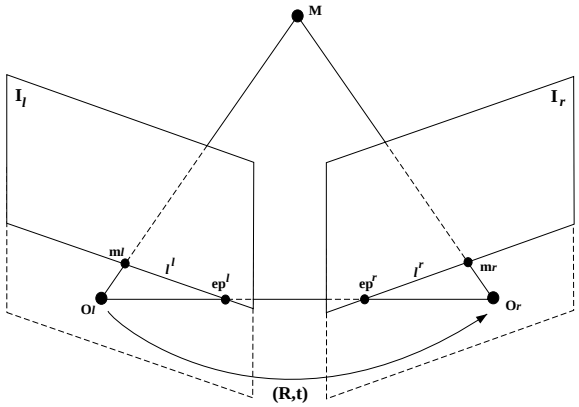

Figure 1: Epipolar constraints associated with two cameras

The left and right camera intrinsic parameters are respectively given by two 3×3 matrices $\mathbf{A}_l$ and $\mathbf{A}_r$. From [15] the essential matrix $\mathbf{E}$ is:

$$\mathbf{E} = \mathbf{T_s R}\,, \tag{1}$$

where $\mathbf{T_s}$ is the skew-symmetric matrix of translational vector $\mathbf{t}$.

The fundamental matrix is given by [7]:

$$\mathbf{F} = \left(\mathbf{A}_l^T\right)^{-1} \mathbf{T_s R A}_r^{-1}\,. \tag{2}$$

Each pair of corresponding points in stereo images satisfies the epipolar constraint

$$m_l^T \mathbf{F} m_r = 0 \tag{3}$$

where m_l and m_r are homogeneous coordinates of a 3D point $\mathbf{M}$ in the scene (Figure 1).

For a pin-hole camera model used here, equation (3) expresses the fact that, when a point $\mathbf{M}$ of the scene is projected on the two image planes, the image points m_l and m_r, the optical centers of the two cameras and the 3D point $\mathbf{M}$ itself lie in a common plane.

The problem of computing the fundamental matrix from points or lines correspondence between two or more images has been largely studied in the past. These techniques can be classified mainly into two groups [18, 9]: linear and non-linear methods. Robust statistical methods were introduced recently for computing the fundamental matrix. They permit outliers rejection before fundamental matrix computation. This improves convergence and leads to more accurate results [18, 16]. Comparisons between different existing methods has shows that non-linear methods can be more accurate that linear ones, but they are computationally

more expensive. It has also been verified that, due to the sensitivity of the methods to matching errors, the computed matrix is often far from the ground truth [18].

2.2 A new linear approach

We have developed a new linear method [1] to compute the fundamental matrix. It has the characteristic of exploiting the available geometrical information about the scene to produce better results. In particular we use the planarity constraints in the epipolar equation (3).

Here is an outline of our approach. Given two homographies (projectif linear mapping) $\mathbf{H}_1$ and $\mathbf{H}_2$ between two corresponding planes Π_1 and Π_2 in two images of a stereo system, we can recover the epipolar line l_r in the right image corresponding to a left image point m_l by:

$$l_r = m_r^1 \otimes m_r^2 = \mathbf{H}_1 m_l \otimes \mathbf{H}_2 m_l$$

The symbol $\otimes$ denotes the cross product.

For n left points we compute their corresponding epipolar lines in the right image plane, these lines must have a common point ep^r called the *epipolar point* [7, 1]:

$$ep^r = l_r(i,j) \otimes l_r(k,l)\ ,$$

where $i,k = 1,\ldots,m; j,l = 1,\ldots,n.\ i = k, j = l$, (m,n) is the image size. The *epipolar point* ep^l in the left image is obtained similarly.

The next step consists of solving a linear system of equations determining the epipolar constraints using singular value decomposition (**SVD**):

$$\begin{bmatrix} -\{\frac{ep^l(2)}{ep^l(1)}\} f_2^T - \{\frac{ep^l(3)}{ep^l(1)}\} f_3^T \\ f_2^T \\ f_3^T \end{bmatrix} m_l = l^r$$

where f_i are the rows of the fundamental matrix. The obtained fundamental matrix satisfies the *rank* 2 constraint [7, 18].

We have conducted several comparisons between different methods for estimating the fundamental matrix. The proposed method shows good results. The computed matrix is close to the ground truth compared to many classical methods, even the non-linear ones. The proposed parametrization does not break down if the epipoles are at infinity. We may deal with that case by selecting a non-zero element from the homogeneous coordinates of the epipole in the denominator. More details on the proposed method and results of comparisons between different techniques of fundamental matrix computation can be found in [1].

3 Geometrically valid point reprojection using the epipolar geometry

In this section we apply the epipolar geometry to the reprojection of the points from the source images to a novel one. In [19], Faugeras and Robert use the knowledge about epipolar constraints inherent to a spatial configuration of three cameras to predict the new positions in the third view of points, lines and curves from their correspondents in the two source images. Figure 2 shows the geometry of these three cameras.

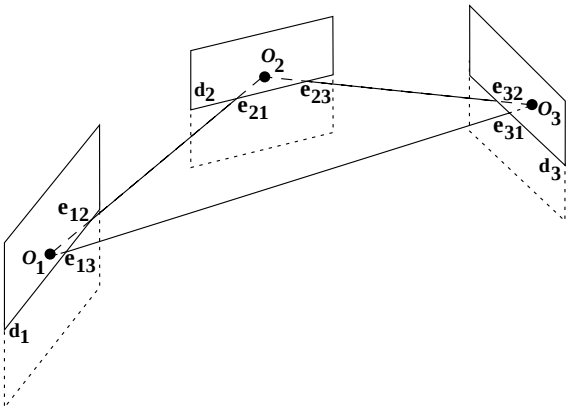

Figure 2: Configuration of three cameras in the scene

Given three cameras denoted by 1, 2 and 3, the geometrical constraints are described by three fundamental matrices $\mathbf{F}_{ij}$ $(i,j = 1,2,3,\ i = j)$. For each image point m_i in the image i, its correspondent epipolar line in the image j is given by $\mathbf{F}_{ij} m_i$. The three optical centers of the cameras define a plane called the *trifocal plane*. The intersection of each image plane with the trifocal plane is denoted by $\mathbf{d}_i$ and contains the epipoles $e_{i,i+1}$ and $e_{i,i+2}$ of camera i with respect to camera $i+1$ and $i+2$ (Figure 2). From the epipolar geometry constraints we have [19]

$$\mathbf{F}_{i,i+1} e_{i,i+2} = \mathbf{d}_{i+1} = e_{i+1,i} \otimes e_{i+1,i+2}\ .$$

Assume m_1 and m_2 are two corresponding points in the images 1 and 2, respectively. The point m_3 in the third view that corresponds to m_1 and m_2, must belong to the epipolar line of m_1 in image 3 given by $\mathbf{F}_{13} m_1$, and to the epipolar line of m_2 in image 3 given by $\mathbf{F}_{23} m_2$. Then m_3 is the intersection of the two epipolar lines $\mathbf{F}_{13} m_1$ and $\mathbf{F}_{23} m_2$ (Figure 3):

$$m_3 = \mathbf{F}_{13} m_1 \otimes \mathbf{F}_{23} m_2\ .$$

The reprojection of lines and curves is also covered by [19], but we will limit ourselves by the transfer of points in this work.

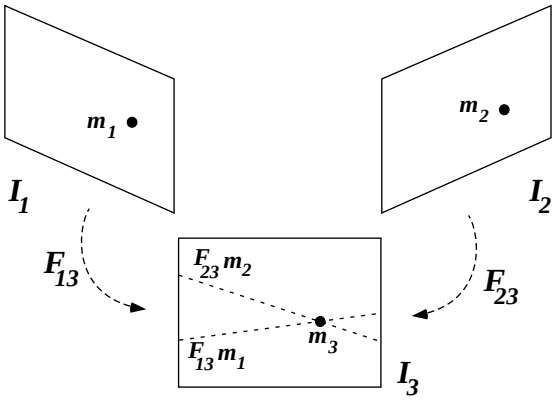

Figure 3: Point reprojection using epipolar geometry

4 Two-dimensional texture mapping

Texture mapping is an important tool in computer graphics 3D modeling. It is used for achieving a realistic rendering of scenes. It consists mainly of a series of spatial transformations. A texture plane is transformed in a 3D surface, and then projected onto the screen. It is achieved by a composition of a mapping function f and a projection P. Here f is a transformation from the texture plane to a 3D surface, and P is the projection from 3D space onto the screen. In texture mapping the 3D objects are usually modeled with planar or bicubic patches for a relevant parameterization between the texture plane and the object space [23, 22, 8, 10].

Image warping deals with 2D geometrical transformations between images. It is based on a spatial transformation: a mapping function that establishes a spatial correspondence between all points in an image and its warped counterpart [23, 8]. It permits a realistic 2D texture mapping. Once a spatial transformation is established from a small set of corresponding points, the interior points are obtained via interpolation. Many geometric transformations techniques exist in the literature. Simple transformations are usually specified by analytic expressions: affine, bilinear, projective, and polynomial. More complex mappings can be estimated from a set of corresponding points defining the interpolation function applied to the remaining interior points. For complex scenes this representation yields to a dense correspondence representation. That implies a point-to-point mapping. In the following we define the projective framework and the *perspective warping* needed in the proposed approach.

A homogeneous spatial 2D transformations can be expressed by a 3×3 matrix $\mathbf{A} = [a_{ij}],\ i, j = 1 \ldots 3$ (the linear transformation of a projective plane onto another projective plane)

$$[x', y', w'] = [x, y, w]\,\mathbf{A} \tag{4}$$

where $[x, y, w]$ are the homogeneous coordinates of an image point in the source image and $[x', y', w']$ are the homogeneous coordinates of its corresponding image point in the warped image.

From $\mathbf{A}$, we can specify a particular mapping function. For example, the affine transformation corresponding to an orthographic (parallel) plane projection is characterized by a transformation matrix with the last column $[0, 0, 1]^T$. In the proposed method we use a perspective mapping function, corresponding to nonzero vector $[a_{13}, a_{23}]^T$.

The perspective transformation is defined up to a scale factor. Without loss of generality, it can be normalized so that $a_{33} = 1$. The perspective transformation is then specified by a projective mapping represented by a 3×3 matrix $\mathbf{T}$:

$$[x', y', w'] = [x, y, 1] \begin{bmatrix} t_{11} & t_{12} & t_{13} \\ t_{21} & t_{22} & t_{23} \\ t_{31} & t_{32} & 1 \end{bmatrix} = [x, y, 1]\,\mathbf{T}$$

Perspective warping is very useful for rendering realistic images when a central projection model is used [23].

A perspective transformation is expressed in terms of nine coefficients, up to a scale factor. Four points correspondence in the source and target images are sufficient to infer this transformation. More points would give more precise results. Once the mapping function (homography mapping) is computed, the remaining points in the new warped image are obtained via a two-dimensional perspective texture mapping. In our approach antialiasing is handled using a linear function.

In the proposed approach we use a perspective warping function to obtain the new image. The mapping function is estimated by the correspondence established between a small set of anchor points in the two images. These anchor points in the target image were reprojected using the epipolar geometry (See Section 3).

5 Experiments

We conducted experiments on images from synthetic and real scenes. The correspondence between points in the source images were selected manually. Then we used epipolar geometry to obtain their positions in each novel view. In particular the fundamental matrix was used to compute the epipolar line corresponding to a given pixel. The intersection of the epipolar lines corresponding to the each pair of corresponding points from

two source images gives the point in the novel view. We have used the method described in section 2 to compute the fundamental matrix. It exploits the existance of the planar patches in the scene. The new image plan was specified using a computed calibration data, in particular the estimated camera intrinsic parameters. The new rotation and translation were selected to give the new fundamental matrix correponding to the desired novel view.

Figure 5 presents the results obtained for a synthetic scene. Two source images (Figure 4) has been taken for our experiment. Here 12 pairs of corresponding points in the two source image were selected. They correspond to scene decomposition into planar surfaces. Then perspective image warping was used to reproject these surfaces in a third view. The warping function was estimated by the *a priori* correspondence between the 12 selected points and their reprojection in the new image plane achieved by mean of epipolar geometry. The remaining image points rendered using perspective image warping appear well-positionned in the synthesized view. We selected features in the synthesized view and compared them with their corresponding features in the real view. An average error obtained is of about 1.8 pixels. The results are close to their real counterpart (See Figure 5.

We used the same system to synthesize the views of a real scene. In this case 14 corresponding points were selected in the source images permitting a decomposition of the scene into planar surfaces. These surfaces were then warped to a novel view using a perspective mapping determined by *a priori* reprojection of the 14 control points. The result is a realistic rendered image, visually close to its real counterpart.

6 Conclusion

In this paper we propose a new approach for novel view synthesis from uncalibrated stereo cameras. It is based on a small number of corresponding points in the reference views. We use the epipolar geometry to produce a geometrically valid reprojection of a small number of reference points in the new image plane. The remaining image points are warped using a perspectively-correct texture mapping between source and novel views.

The proposed approach permits an efficient way for view synthesis of polygonal objects. Complex scenes are handled using a multiresolution image synthesis scheme similar to that in computer graphics 3D modeling. Starting with a rough decomposition and synthesis of the novel image. The result can be refined in subregions by a finer decomposition of the image in each subregion, up to a point-to-point reprojection in complex scenes where high level of details is important.

The results presented here show that the proposed method gives a good alternative to classical image-based rendering techniques, without the computational burden of finding dense correspondence between the source images or a priori 3D reconstruction of the scene.

The proposed approach can be used in different applications where fast rendering is needed, for example, videoconferencing, immersive telepresence, realistic virtual worlds. The use of the method in telerobotics is presented in [2].

References

[1] M.A. Akhloufi, W.B. Tong, V. Polotski, and P. Cohen. Estimating the fundamental matrix for a stereoscopic system from planar surfaces. In *Proc. Fourth Joint Conference on Information Sciences, Durham (NC)*, 1998.

[2] P. Cohen, J.-Y. Hervé, and M.A. Akhloufi. Augmented reality concepts for mining vehicle operation. In *CIM/CMMI/MIGA Conf., Montreal*, 1998.

[3] S. Avidan and A. Shashua. Novel view synthesis in tensor space. In *Conference on Computer Vision and Pattern Recognition*, pages 1034–1040, 1997.

[4] S. Ullman and R. Basri. Recognition by linear combinations of models. *IEEE. Trans. on Pattern Analysis and Machine Intelligence*, 13(10):992–1006, 1991.

[5] S.E. Chen and L. Williams. View interpolation for image synthesis. In *Proc. Siggraph 93*, pages 279–288, 1993.

[6] P.E. Debevec, C.J. Taylor, and J. Malik. Modeling and rendering architecture from photographs. In *Proc. Siggraph 96*, pages 11–20, 1996.

[7] O. Faugeras. *Three-Dimensional Computer Vision: a Geometric Viewpoint.* MIT Press, 1993.

[8] J.D. Foley, A. Van Dam, S.K. Feiner, and J.F. Hughes. *Computer Graphics: Principles and Practice.* Addison-Wesley, 2nd edition, 1990.

[9] R.I. Hartley. In defence of the eight-point algorithm. *IEEE. Trans. on Pattern Analysis ans Machine Intelligence*, 19(6):580–593, 1997.

[10] P. Heckbert. Survey of texture mapping. *IEEE. Computer Graphics and Applications*, 6(11):56–67, 1986.

[11] T. Huang and A. Netravali. Motion and structure from feature correspondences: A review. In *Proc. IEEE*, volume 82, pages 252–268, 1994.

[12] S. Moezzi, A. Katkere, D.Y. Kuramura, and R Jain. Reality modeling and visualization from multiple video sequences. *IEEE. Computer Graphics and Applications, Nov.*, pages 58–63, 1996.

[13] T. Kanade, P. Rander, and Narayanan. Virtualized reality: Constructing virtual worlds from real scenes. *IEEE. Multimedia*, 4(1):34–47, 1996.

[14] K.N. Kutulakos and J. Vallino. Affine object representation for calibration-free augmented reality. In *Proc. of Virtual Reality International Symposium*, pages 25–36, 1996.

[15] H. Longuet-Higgins. A computer algorithm for reconstructing a scene from two projections. *Nature*, 293:133–135, 1981.

[16] Q.-T. Luong and O.D. Faugeras. The fundamental matrix: Theory, algorithms and stability analysis. *The International Journal of Computer Vision*, 17:43–76, 1996.

[17] L. McMillan and G. Bishop. Plenoptic modeling: An image-based rendering system. In *Proc. Siggraph 95*, pages 39–46, 1995.

[18] P.H.S. Torr and D.W. Murray. The development and comparison of robust methods for estimating the fundamental matrix. *The International Journal of Computer Vision*, 24:271–300, 1997.

[19] O. Faugeras and L. Robert. What can two images tell us about a third one? In *Proc. Third European Conf. on Computer Vision*, pages 485–492, 1994.

[20] S.M. Seitz and C.M. Dyer. View morphing. In *Proc. Siggraph 96*, pages 21–30, 1996.

[21] A. Shashua. Projective structure from uncalibrated images: structure from motion and recognition. *IEEE. Trans. on Pattern Analysis ans Machine Intelligence*, 16(8):778–790, 1994.

[22] A. Watt. *3D Computer Graphics*. Addison-Wesley, 1993.

[23] G. Wolberg. *Digital Image Warping*. IEEE Computer Society Press Monograph, 1990.

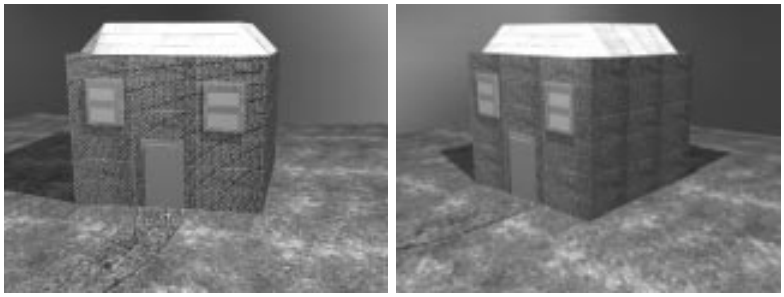

Figure 4: Source views from a house sequence

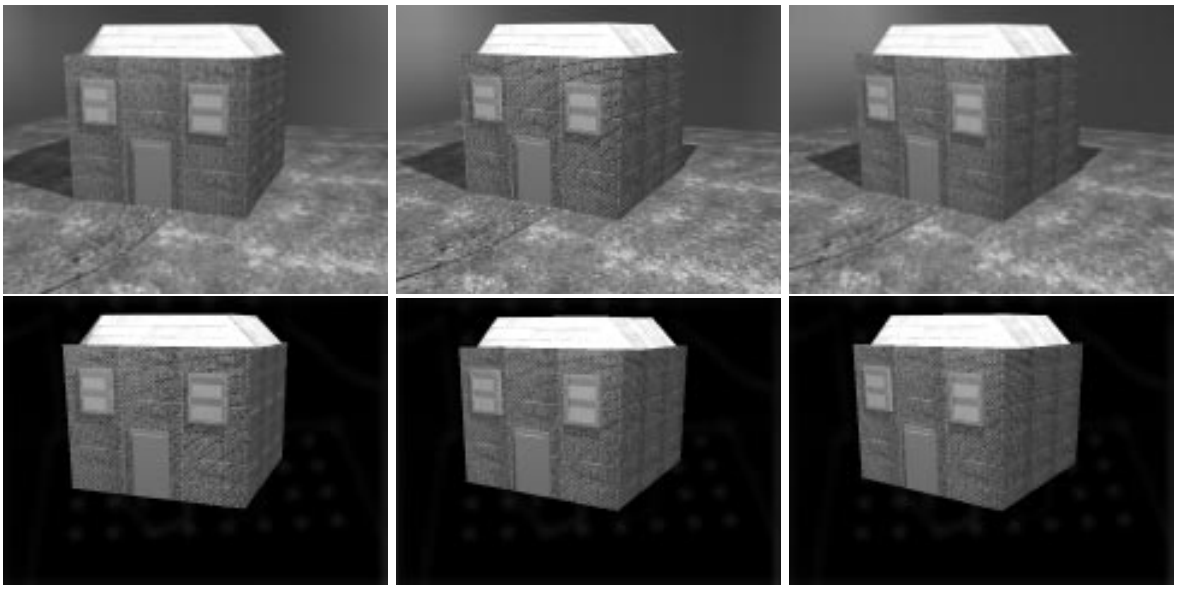

Figure 5: Resulting synthesized views(down) and their corresponding real images(up)