# View Prediction for Improved Visual Servoing

A.H. Abdul Hafez
CSE Department, College of Engineering,
Osmania University, Hyderabad-07, India.
hafezsyr@ieee.org

Piyush Janawadkar
Center for Visual Information Technology,
IIIT, Hyderabad-19, India.
piyush_j@students.iiit.ac.in

C.V. Jawahar
Center for Visual Information Technology,
IIIT, Hyderabad-19, India.
jawahar@iiit.ac.in

*Abstract* — **Visual Servoing is an important problem in robotic vision. In this paper we propose algorithms for visual servoing using novel view prediction. A visual servoing control law robust to large proportion of noisy outliers in image data is proposed. The method employs robust statistical techniques and novel view prediction. We also propose a visual servoing control law which results in faster convergence into a desired position. This algorithm employs novel view synthesis to minimize an error over a number of future views. We also show that view prediction can be effectively used for retaining the image features in the camera field of view during the motion. These techniques are validated with experiments and compared with other robust methods in a simulation framework.**

## I. Introduction

Visual servoing is the process of positioning the end effector of a robot manipulator with respect to a target object or a set of features. This is achieved by processing the visual feedback and minimizing an appropriate error function. The visual feedback can be image features or object pose with respect to the camera frame.

Based on the visual information, visual servoing systems can be classified to three categories [1]: position-based (3D), image-based (2D), or hybrid ($2\frac{1}{2}$D) visual servoing. In image based visual servoing, 2D visual information is extracted from the image and used directly in the control law to generate the control signal *i.e.*, the screw velocity of the robot end-effector. This velocity is computed from an error function of the image space features and through the image Jacobian or the interaction matrix [2]. The accuracy of the computation depends on features detection, matching, tracking, and modeling algorithms employed during the process. If the correspondences between features contain errors, the visual servoing process fails to converge, and the system will reach inaccurate final state or a local minima [3].

In addition to the convergence issues, features may get out of the camera field of view and visual servoing can become intractable. Without any constraints in the image space, as in the case of position based-based visual servoing, the image features may leave the camera field of view. Features may leave the camera field of view when there is a rotation about $Z$ axis for features near to the i view boundaries [4]. It could also happen in presence of a constant interaction matrix estimation [5].

This paper presents three applications of novel view synthesis to the visual servoing problem.

- We show that a visual servoing control law robust to image noise and measurements error can be derived with the help of view prediction.
- Another application of view synthesis is in minimizing the error function in an image based visual servoing algorithm based on information available from future views. This results in faster convergence.
- Finally we demonstrate its application in retaining the image features in the camera field of view during the servoing process.

In our work [6], we proposed a new method for robust visual servoing using multiple view geometry. We employ the epipolar constraints to produce an image based visual servoing control law that is robust to image noise. In the visual servoing literature, Comport *et al.* [7] proposed an M-estimator based statistical approach that utilizes redundancy in image features to detect and reject the outliers. The inability to reject outliers in presence of excessive noise is a drawback of this method. Our method uses both epipolar geometry and statistical techniques for robust visual servoing. Image-based visual servoing needs initial and desired images for calculation of the motion parameters. We improve the robustness by using an additional image with known relationship to the initial frame. From the image acquired by the camera and a predicted image from the two known frames, we identify outliers for improving the robustness.

View prediction can also help in improving the convergence time and keeping the image features in the camera field of view. Novel view synthesis makes it possible to minimize the error function along a future horizon of iterations. Classical methods minimize the error measure computed from one (current) set of image measurements. The velocity imparted to the robot arm at every time instant, is used to compute the expected pose in future time instances. Predicting the next view will give a sense about the next position of the feature in the image. If the feature is going to get out of the camera field of view, a backward translation along the $Z$ of the camera frame is applied to keep the feature within the boundary of the view.

## II. Background

### A. Image-based Visual Servoing

The problem of image-based visual servoing is that of positioning the end-effector of a robot arm such that a set of image features $S$ reaches a desired value $S^*$. The set $S$ can be composed of the coordinates of points that belong to the target object. Other kind of geometric features like straight line segments, angles, or spheres can also be used. Consider the error function

$$e(S) = S - S^*, \tag{1}$$

which is the difference between the current feature vector $S$ and the desired one $S^*$. By differentiating this error function with respect to time, with the desired features $S^*$ remaining constant, we get

$$\frac{de}{dt} = \frac{dS}{dt} = (\frac{\partial S}{\partial P})\frac{dP}{dt} = L_S V, \tag{2}$$

where $S$ is a $(2N \times 1)$ features vector represented by the image coordinates $(x, y)$, and $N$ is the number of points. The velocity $V = (v^T, \omega^T)^T$ is the camera velocity, $v$ is translational velocity and $\omega$ is rotational velocity. The pose vector $P = (x, y, z, \alpha, \beta, \gamma)$ is a $(6 \times 1)$ vector, where $(x, y, z)$ represent the 3D coordinates of the camera frame position and the three angles $(\alpha, \beta, \gamma)$ represent the camera frame direction with respect to a reference frame. The $(2N \times 6)$ matrix $L_S$ is called the interaction matrix or the image Jacobian. It relates the changes in the image space to the changes in the Cartesian space.

The main objective of the visual servoing process is to minimize the error function $e(S)$. For exponential convergence of the minimization process , we need $\frac{de(S)}{dt} = -\lambda e(S)$. By substituting in (2) and using a simple proportional control law, the required velocity of the camera can be shown to be

$$V = -\lambda L_S^+ e(S), \tag{3}$$

where $L_S^+$ is the pseudo inverse of the Jacobian matrix $L_S$, and $\lambda$ is a scale factor.

A robust visual servoing control law based on M-estimator was proposed in [7]. The error function was modified to be

$$e(S) = D[S - S^*],$$

where $D = diag(w_1, .., w_i, .., w_{2N})$ is a weighting matrix, and $N$ is the number of points. The weight $w_i = 0$ if the point is an outlier and $w_i = 1$ if the point is an inlier. The computation of weights $w_i$ is done using Tukey's robust function [8]. For this objective function, the control law is given as

$$V = -\lambda[DL_S]^+ D[S - S^*]. \tag{4}$$

One can see that the matrix $D$ is being introduced to the error function and the interaction matrix. Entries of the interaction matrix that correspond to the outlier features also will be nullified by the multiplication of zeros.

### B. Novel View Synthesis

Epipolar Geometry describes the relationship between corresponding points in two views. Suppose $x$ and $x'$ are the corresponding points in two views, then the epipolar constraint has the form [9]

$$x^T E x' = 0. \tag{5}$$

The $(3 \times 3)$ matrix $E$ is known as the *Essential matrix* and has the rank 2. The Essential matrix maps a point $x'$ in one view to a line $l = Ex'$ in the other view. This line is called the epipolar line. The essential matrix depends on the relative geometry of the two cameras. Suppose the relative transformation between the two cameras is given by $T = \left[\begin{array}{cc} R & \mathbf{t} \end{array}\right]$ then, it can be shown that

$$E = R^T [t]_\times, \tag{6}$$

where the matrix $[t]_\times$ is the antisymmetric matrix associated to vector $\mathbf{t}$. The pairwise epipolar geometry can be used to predict new views [9]. A correspondence between two given images $(x \leftrightarrow x')$ constrains the point in the third image $x''$ to lie on the lines $E_{31}x$ and $E_{32}x'$. The point in the third view is the intersection of these two epipolar lines and is given by

$$x'' = E_{31}x \times E_{32}x'. \tag{7}$$

From the correspondence between two original views and the required view defined in terms of essential matrices, a novel view can be produced by transferring all corresponding pixels to the new view by this method.

## III. The Proposed Solutions

### A. Image-based Visual Servoing Using Predictive views

It is well known that given two views of a scene, a third view from any desired location can be computed. Such view prediction techniques may be used in conjunction with visual servoing to achieve faster convergence. The basic idea is to use view prediction to predict errors in $N$ views that would be obtained by moving $N$ steps as per the classical control law, and then to compute the optimal velocity for faster convergence based on this additional information. Our approach is as follows.

Let us choose the following objective function

$$\mathcal{E}_i = \frac{1}{2}\sum_{j=i}^{i+N}(S_j - S^*)^T(S_j - S^*), \quad (8)$$

where $\mathcal{E}_i$ is the expected error in the next $N$ time intervals. Since, $\forall j > i, S_j$ is a function of $S_i$, we minimize $\mathcal{E}_i$ as a function of $S_i$. We proceed by iteratively minimizing $\mathcal{E}_i$ using Gradient Descent. At each step, we update feature vector $s_i$ in the following manner

$$S_{i+1} = S_i - \eta\frac{\partial \mathcal{E}_i}{\partial S_i}. \quad (9)$$

We compute the gradient as

$$\nabla\mathcal{E}_i = \frac{\partial \mathcal{E}_i}{\partial s_i} = \sum_{j=i}^{i+N}\left[(S_j - S^*)^T\left(\frac{\partial S_j}{\partial S_i}\right)\right]^T. \quad (10)$$

The term $\frac{\partial S_j}{\partial S_i}$ can be written as

$$\frac{\partial S_j}{\partial S_i} = \frac{\partial S_j}{\partial r_j}\frac{\partial r_j}{\partial S_i} = \frac{\partial S_j}{\partial r_j}\frac{\partial r_j}{\partial t}\left(\frac{\partial S_i}{\partial t}\right)^+ = L_jV_j(L_iV_i)^+ \quad (11)$$

From equations (9 ), (10), and (11) and by some simplification, we obtain the generalized control law the generalized control law

$$\hat{V}_i = -\frac{\eta}{\Delta t}L_i^+(L_iV_i)^{+T}\left[\sum_{j=i}^{i+N}(S_j - S^*)^T L_jV_j\right]^T, \quad (12)$$

where $V_i$ and $L_i$ are the velocity vector and the Jacobian matrix estimated from the current view. The velocity vector estimated along the $N$ views starting from the current view $i$ is $\hat{V}_i$.

### *B. Robust Image-based visual servoing based on predictive views*

Here we propose a solution to the case of the proportion of the noisy points is more than 50%. An additional reference image is used to predict a virtual novel image. The transformations between the current image and each of the two initial and reference images are used to predict the novel image. The predicted image is used to identify the outlier points in the current image. A classification process is defined to use the error between the data points in the current image and its correspondences in the predicted image as a discriminant function. During the visual servoing process, a constant value is assigned to the error function which corresponds to the outlier point feature. The error function given in (1) is modified as

$$\hat{e}(S_i) = \begin{cases} S_i - S^* & \text{if } S_i \text{ is inlier.} \\ e_0(S_i) & \text{if } S_i \text{ is outlier.} \end{cases}, \quad (13)$$

here $e_0(S_i)$ is a precomputed constant value of the visual servoing error function. By substituting the error function given in (13) in the equation (3), the control law will be modified to becomes

$$\hat{V}_i = -\lambda L_i^+\hat{e}(S_i), \quad (14)$$

Consider an initial image $I^0$ of a scene consisting of a set of 3D points. In addition to the initial image, the camera takes another image (reference image) $I^r$ with a known transformation between these initial and reference camera positions $T_{0r} = [R_{0r}, t_{0r}]$. Select a set of point features $S^0$ in the initial image $I^0$ and another corresponding set $S^r$ in the reference image $I^r$. The novel image contains the corresponding set to these two sets of point features.

The novel image computation is done using the velocity measurement of the camera. At each iteration of the visual servoing process, the transformations between the current image and the two (initial and reference) images $T_{0i} = [R_{0i}, t_{0i}]$ and $T_{ri} = [R_{ri}, t_{ri}]$ are computed. By substituting these two transformations in (6), the essential matrices $E_{0i}$ and $E_{ri}$ are obtained. Using these essential matrices and equation (7), the current predicted image is computed.

The discriminant error function is defined. The set of features $\hat{S}^i$ in the current predicted image is corresponding to the sets $S^0$ and $S^r$. Substituting the set $\hat{S}^i$ in the control law which is given by (3) will give the camera velocity that has been contributed by the $\hat{S}^i$ data set

$$V = -\lambda L_{\hat{S}^i}^+(\hat{S}^i - S^*). \quad (15)$$

Consider the term $r_{pred} = \lambda(\hat{S}^i - S^*)$ as the state of the current predicted image with respect to the desired one, and the term $r_{actu} = \lambda(S^i - S^*)$ as the state of the actual or measured current image with respect to the desired one. The error required for the discriminant function is defined as $r_i^2 = (r_{pred} - r_{actu})^2$. In literature, these are known as the residual values of the points $S^i$. Using (2), (3) and (15), we can write this error for each single feature $S_i$ in the actual current image as

$$r_i^2 = (L_{S_i}V + \lambda(S_i - S_i^*))^2. \quad (16)$$

The point $S_i$ is classified as an outlier if $r_i \geq t_\sigma$, and is classified as an inlier if $r_i < t_\sigma$, where the threshold value $t_\sigma = \sqrt{5.99}\sigma$ [10].

### *C. Keeping the Features Visible*

In order to keep all image features visible in the camera field of view at all times, a backward translation along the $z$ axis is applied only when an image point estimation is close enough from the image boundaries. Since a translation in the negative $z$ direction of the camera frame will cause image point to move toward the center of the image, The distance between the the points and the boundary of the image is mapped directly to the $V_z$ components of the velocity computed by the control law.

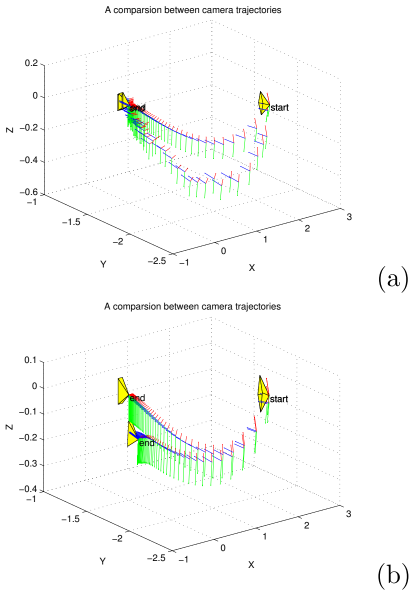
$$V_z = f(d(u_i, v_i)), \tag{17}$$

where $i = 1, .., N$ and $d(u_i, v_i)$ is the smallest distance of the point $(u_i, v_i)$ to the image boundary in the next estimated view, $N$ is the number of point considered as features. In addition, to keep tracked estimation of the next view, we need always to store the previous view and the transformation from it to the current view. By computing the velocity from the current view, predicting the next view is straight forward using the essential matrices as explained in section II-D. Now we look for the smallest distance of each point to the boundaries of the predicted view. If $d(u_i, v_i) \leq d_0$, then $V_z = -a$, where $a$and $d_0$ are a suitable constant values. If the value of $a$ is too small, the point movement toward the image center will not be enough to keep the point feature visible in the view. In contrast, if the value of $a$ is very high, it will push the point much far toward the image center and the trajectory will be undesirable. The value of the another constant $d_0$ should be suitable to the uncertainty and delay in the robot arm and camera dynamics.

## IV. Experimental Results

In the simulation experiments we used a set of 3D points $X_i$, $i = 1, ..., N$. These points belong to an object in the scene. A positioning task is considered for the study. The robot arm has to move from initial position to a desired position given as a desired image of the object. We conduct our simulation experiments in presence of excessive noise using two robustness methods. The noise was introduced in the matching and feature extraction step. The image point coordinates are considered as features. Since we have $N$ points, the total number of features is $2N$. In other words, features $S_{2i}$ and $S_{2i-1}$ are produced by the point $x_i$. If $S_{2i}$ or $S_{2i-1}$ is considered as an outlier, other one also labeled as an outlier also. The error given in Equation (13) will be used for both the features.

In the first experiment a considerable noise was added to 8 points out of 12 points. Without the proposed method, (as shown in Figure 1 (b)) configuration leads to an unstable situation or a local minima. Final state is much different from the desired one. Better results are obtained in the M-estimator based control law case reported in [7], but the final state is still different from the desired one. As shown in Figure 1 (c), M-estimator based method is not able to detect the outliers in the case of large proportion ( in our example 8 points out of 12). The third case is where we use our proposed method. The method was tested on different noise levels and proportions of outliers. Figure 1 (d), shows the result of our novel view based method in the case of large amount of noise was added to 8 points. They show that the final state has reached the exact desired position, and the error perfectly converged to the minimum value. Figure 2 shows a comparison between the camera trajectories in the Cartesian space in the ideal case and each of M-estimator method in Figure 2 (b) and our proposed method in Figure 2 (a). The difference between the final and desired states is clear. From this experiment we can conclude that our method is efficient regardless to the number of image points disturbed by noise. M-estimator methods are restricted to the case where less than 50% of the points were distrubed by noise.



(a)

(b)

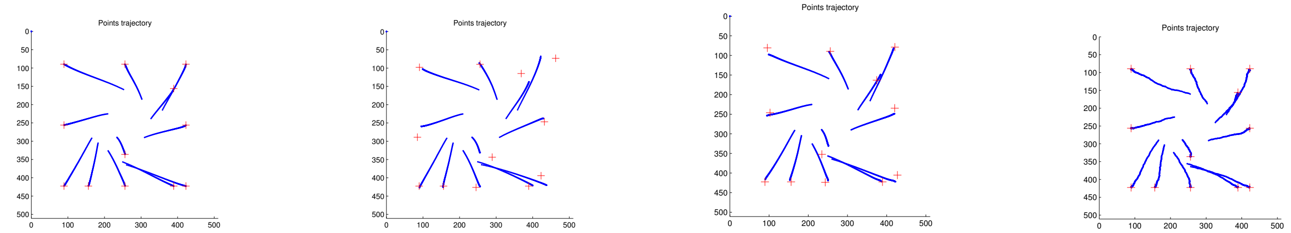
Fig. 2. Camera trajectory comparison with the reference case. (a) with the M-estimator, (b) with the novel view method.

The second experiment was done using the control law given by equation 12 in section III-A. This control law is developed based on the prediction of the future views. Figure(3) depicts a comparison between the feature coordinates error and trajectory of features in the image space, this is in the two predictive view error and classical visual servoing error. It may be observed that the proposed method converges faster.

The third experiment was curried out using the control law proposed in [4] incorporated with the method presented in section III-C to keep the features in the camera field of view. The results are depicted in Figure(4). Originally, the features are supposed to move along a circle. Because of the effects of our proposed method to keep the features visible, it moved along the shown trajectory.

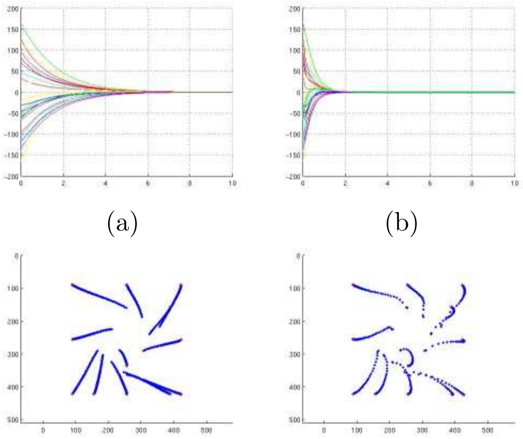
## V. Conclusion

A novel method has been proposed here to give a solutions to visual servoing problems. One solution is a robust



(a) (b) (c) (d)

Fig. 1. Points trajectory in the image space of image features in case of the ideal scenario (a), noisy features without robustness (b), using M-estimator for robustness (c) using our proposed novel view for robustness (d). The mark + in the image indicates the desired position of the image point.



(a) (b)

(c) (d)

Fig. 3. A comparison in the convergence time and image trajectory between predictive view visual servoing (b,d)and classical visual servoing (a,c).

image-based visual servoing control law. This method classifies the data points to outliers or inliers. The detected outlier is introduced in the control law with a constant error value. The core of this method lies in combining between statistical methods and multiple view geometry. As an improvement to the previous work in the robust visual servoing, this method can deal with large noisy features proportion, even more than 50%. Another solution is a fast convergence control law based on novel view synthesis. In this control law an error function has been minimized along a set of future set of views instead of minimizing along tyhe current view as in the classical visual servoing. The last solution is to the problem of leaving the image features the camera field of view. Using the stimation of the next predictive view a reaction is considered to mintain the feature in the visible part of the image. As a future work, this can be extended to visual servoing architectures like 3D and $2\frac{1}{2}D$ visual servoing. Other kind of featurs may be considered to improve the robustness.

## References

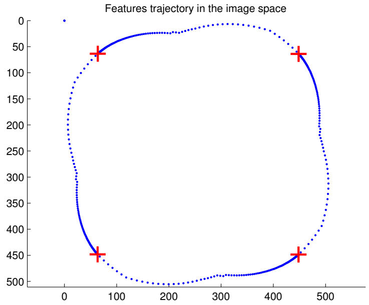


Fig. 4. Features trajectory in the image space for the predictive view based keeping features in the camera field of view method

[1] E. Malis, F. Chaumette, and S. Boudet, $2\frac{1}{2}D$ visual servoing, *IEEE Trans. on Robotics and Automation* , Vol. 15, pp. 234-246, 1999.

[2] S. Hutchinson, G. Hager, and P. Cork, A Tutorial on visual servo control, *IEEE Trans. on Robotics and Automation* , Vol. 17, pp. 18-27, 1996.

[3] E. Marchand, and F. Chaumette, Feature tracking for visual servoing purposes, *Robotics and Autonomous Systems*, Vol. 52, pp. 53-70, 2005.

[4] P. Corke, and S. A. Hutchinson, A New Partitioned Approach to Image-based Visual Servo Control, *IEEE Trans. on Robotics and Automation* , Vol. 17, No. 4, pp. 507-515, 2001.

[5] F. Chaumette, Potential problems of stability and convergence in image-based and position-based visual servoing,*The Confluence of Vision and Control, D. Kriegman, and G. Hager, pp. 66-78, No 237, Springer-Verlag, 1998*

[6] A.H. Abdul Hafez, Piyush Janawadkar, and C.V. Jawahar, Robust Visual Servoing Based on Novel View Prediction, *Int. Conf. on Computational Intelligence, Robotics and Autonomous Systems, CIRAS'05* , Singapore, Dec, 2005.

[7] A.I. Comport, M. Pressigout, E. Marchand, F. Chaumette, A Visual Servoing Control Law that is Robust to Image Outliers, *IEEE Int. Conf. on Intelligent Robots and Systems, IROS'03* Vol. 1, pp. 492-497, Nevada, 2003.

[8] P. J. Huber, *Robust Statistics*, Wiler. New York, 1981.

[9] Richard Hartley and Andrew Zisserman, *Multiple View Geometry in Computer Vision*, Cambridge University Press, 2003.

[10] P.J. Rousseeuw, A.M. Lero, *Robust Regression and Outlier Detection*, John Wiley and Sons, 1987.