A ROBUST ALGORITHM FOR FUSING NOISY DEPTH ESTIMATES USING STOCHASTIC APPROXIMATION

Amit K. Roy Chowdhury, R. Chellappa

Center for Automation Research
Department of Electrical and Computer Engineering
University of Maryland, College Park
MD 20742, USA
{amitrc, chella}@cfar.umd.edu

ABSTRACT

The problem of structure from motion (SFM) is to extract the three-dimensional model of a moving scene from a sequence of images. Most of the algorithms which work by fusing the two-frame depth estimates (observations) assume an underlying statistical model for the observations and do not evaluate the quality of the individual observations. However, in real scenarios, it is often difficult to justify the statistical assumptions. Also, outliers are present in any observation sequence and need to be identified and removed from the fusion algorithm. In this paper, we present a recursive fusion algorithm using Robbins-Monro stochastic approximation (RMSA) which takes care of both these problems to provide an estimate of the real depth of the scene point. The estimate converges to the true value asymptotically. We also propose a method to evaluate the importance of the successive observations by computing the Fisher information (FI) recursively. Though we apply our algorithm in the SFM problem by modeling of human face, it can be easily adopted to other data fusion applications.

1. INTRODUCTION

The problem of structure from motion (SFM) is to extract the three-dimensional model of a moving scene from a sequence of images. Traditional SFM algorithms [1], [2] recover a 3D scene structure from two images. However, these algorithms often produce inaccurate reconstructions of the scene, mainly due to incorrect estimation of camera motion. Recently techniques have been developed that use multiple images for scene reconstruction, achieving greater robustness and accuracy by fusing the two-frame estimates [3], [4], [5], [6].

2. OVERVIEW OF THE ALGORITHM

2.1. The Observation Model

Let $\mathbf{d}_i$ represent the depth value obtained by the two-frame SFM algorithm from the $i$ and $(i+1)$-th frame and $\mathbf{d}_i(n)$, the $n$-th position in that vector. We assume a linear observation model

$$\mathbf{d}_i = \mathbf{H}u + \mathbf{v}_i, \quad i = 1, ..., K$$

where $\mathbf{v}_i$ is a noise process with unknown distribution. $\mathbf{H}$ is assumed diagonal [4] implying that the depth at every point

This paper describes a new data fusion algorithm applied to multi-frame structure from motion (MFSFM) using stochastic approximation (SA). The method can be easily extended to other data fusion applications also. We assume that the 2-frame estimates (observations) are available from a suitable 2-frame SFM algorithm. The correspondence problem is not addressed in this paper.

We propose a recursive strategy for estimating the unknown true depth given the observations. Our observation model assumes additive noise on the true value but does not assume any particular distribution of the noise process. We also take care of eliminating outliers in the observation sequence by choosing an appropriate cost function. Our algorithm uses the RMSA technique. We show that the estimates converge to the true value asymptotically. We also propose a method for evaluating the importance of the successive observations (i.e. the number of frames to consider) by evaluating the Fisher Information (FI) recursively. The results of our algorithm are demonstrated by applying them on image sequences of a human face.
For Gaussian random variables, all odd central moments are identically zero (from symmetry) and all cumulants of order greater than two are zero. However, as seen from Fig. 1, the observations need not follow a normal distribution.

An analysis of the depth values across frames for any pixel shows that there are some values which can be characterized as outliers and should not be considered while fusing the estimates (see Fig. 2).

It is a well-known fact that the median is less sensitive to outlying data points than the mean. The mean of a sample of size $n$ has breakdown $1/n$ since by changing just one data value we can force the sample mean to have any value whatsoever. The sample median has breakdown 50% (the best possible value), reflecting the fact that it is less sensitive to individual values. The least squares regression estimator inherits the sensitivity of the mean and has breakdown $1/n$, whereas the least median of squares estimator (LMS) has breakdown roughly 50%. Thus the cost function we try to optimize is

$$ u^* = \arg \min_u \left( \text{median}(d_i - u)^2 \right) \quad (3) $$

The disadvantage of this method is that we no longer have a closed form solution for the optimal estimate as we had for the mean-square error criterion.

### 3. A Recursive Algorithm Using Stochastic Approximation

#### 3.1. The Robbins-Monro Algorithm

The Robbins-Monro stochastic approximation algorithm is a stochastic search technique for finding the root $\theta^*$ to $g(\theta) = 0$ based on noisy measurements of $g(\theta)$, i.e. $Y_k(\theta) = g(\theta) + c_k(\theta), k = 1, \ldots, K$, where $c_k(\theta)$ is assumed to be the noise term and $K$ is the number of observations. The RMSA algorithm obtains the estimate by the following recursion,

$$ \hat{\theta}_{k+1} = \hat{\theta}_k - a_k Y_k(\hat{\theta}_k). \quad (4) $$

where $a_k$ is an appropriately chosen sequence. Details of the algorithm can be found in [7], [8]. We will outline the method for obtaining the solution for our specific problem. Suppose that $F_X(x)$ is the unknown distribution of a sequence of observations $X_0, X_1, \ldots$ and we are interested in finding the root of the equation $g(\theta) = F_X(\theta) - 0.5 = 0$, i.e. the median of the distribution. For this problem, the Robbins-Monro (RM) recursion is as follows [8]:

$$ \hat{\theta}_{k+1} = \hat{\theta}_k - a_k s_k(\hat{\theta}_k) - 0.5 \quad (5) $$

where

$$ s_k(\hat{\theta}_k) = \begin{cases} 1 & \text{if } X_k \leq \hat{\theta}_k \\ 0 & \text{otherwise} \end{cases} \quad (6) $$
The choice of the gain sequence \( a_k \) is determined by the convergence properties of the algorithm \cite{9}, \cite{8}. 2

The sequence in consideration in our case is \( X_i(u) = \overline{(d_i - u)^2} \). The minimization is carried out over a predetermined search set \( \mathcal{U} \) and the number of frames is determined by analyzing the Fisher information of the observations. Since the depth observations \( \{d_i\} \) are the result of a 2-frame SFM algorithm, they are corrupted by noise whose distribution is unknown in general. However, since RMSA solves for the distribution operator, we obtain the median of the sequence using RMSA for the particular problem. From the above we see that given the sequence of observations \( \{X_i(u)\} \), we wish to obtain the median of the sequence using RMSA for the particular value of the parameter \( u \). The estimate of the median obtained by the RMSA recursion is strongly consistent. Also, the error in the estimate converges in distribution to a normal variable with a density \( f_X \), we can write

\[
\frac{d}{d\theta} \log f_Y(y) = \frac{d}{d\theta} \log f_X(y - \theta) = \frac{d}{dt} \log f_X(t) \frac{dt}{d\theta}, \quad t = y - \theta
\]

The estimate of the gradient of \( y(\theta) \) with respect to \( \theta \in \mathbb{R}^p \):

\[
\hat{g}(\theta) = \frac{y(\theta + \Delta) - y(\theta - \Delta)}{2} \begin{bmatrix} \Delta^{-1}_{1} \\ \vdots \\ \Delta^{-1}_{p} \end{bmatrix}
\]

where \( \Delta = (\Delta_1, \ldots, \Delta_p) \) and the components of \( \Delta \) are independent Bernoulli random variables. The steps in computing the Fisher information are:

**Step 1** Given \( \theta_k \), generate a set of \( k \) pseudo measurements according to the empirical distribution of the observations. Denote these by \( x_{\theta_k} = x_{\theta_k}(k) \). Calculate the gradient according to (8). It may be necessary to average several gradient estimates with independent values of \( \Delta \). Compute the term within the expectation operator in the definition of Fisher information (7).

**Step 2** Repeat Step 1 a large number of times, say \( N \). Average the estimates obtained. This is the estimate of the Fisher information, \( F_k(\theta_k) \).

We can evaluate the relative importance of the observations by looking at increase in the Fisher information (see Fig. 3).

### 4. RESULTS AND ANALYSIS

We applied our algorithm for 3D modeling of human faces from 2D images. Given a sequence of images, we used the two frame algorithm described in \cite{2} to obtain the depth map. In this method, a fast partial search is used to compute the motion and structure. The least squares error of the system is computed using Fourier techniques and the focus of expansion is estimated in \( O(N^2 \log N) \) operations for a \( N \times N \) flow field. The two-frame depths were then fused by the method described above. A 3D model was created by interpolating the values at the pixels at which the depth was not obtained. From this model, we synthesized views which are

---

2 We used the commonly chosen gain sequence \( a_k = 0.1/(k+1)^{101} \).

3 \( E_\theta \) represents expectation with respect to \( \theta \) and \( \nabla_\theta \) represents the gradient with respect to \( \theta \).
We have demonstrated the optimality of the method by showing that the estimate is strongly consistent and asymptotically normal. We have also presented a method for computing the Fisher information using stochastic gradient and applied it to compute the number of frames to consider for the fusion algorithm. The work was applied to the modeling of human faces and the results have been presented.

6. REFERENCES