Stereo Reconstruction Using HMM and Particle Filtering

Sanja Damjanović, Ferdi van der Heijden and Luuk J. Spreeuwers

***Abstract*—We investigate a new approach to stereo reconstruction using probabilistic techniques and demonstrate that particle filtering is a suitable technique for this application. The advantage of particle filtering over other approaches is its flexibility and ease to include more complex knowledge of the scene into the probabilistic model. We perform the reconstruction of the 3D structure from a pair of rectified stereo images assuming that the scene statistics is described by the first order hidden Markov model (HMM). The stereo matching is treated as the state estimation where the state variable is the disparity. Evolution of the state variable happens along the epipolar line. The transition probabilities allow for continuous and abrupt transitions, i.e. changes in disparity. The likelihood values are derived using the normalized crosscorrelation map (NCC).**

This paper presents the first implementation of particle filtering and HMM applied to stereo reconstruction. We demonstrate that particle filtering can be successfully applied in 3D reconstruction.

***Index Terms*—Stereo reconstruction, HMM, probabilistic framework, particle filtering and smoothing**

I. Introduction

We approach the stereo matching problem as the state estimation problem [1]. We choose the prior model as the 1D HMM and the likelihood values derived on the base of the NCC coefficients and apply different probabilistic algorithms for disparity calculation: the forward, the forward/backward and Viterbi algorithm. A suitable algorithm for the state estimation under the conditions present in this case of the stereo matching is the particle filtering algorithm and the particle filtering followed by smoothing, [2], [1]. The disparity maps calculated by the probabilistic algorithms are compared with the dynamic programming result in which the cost function is derived from the NCC coefficients. This approach has up to now never been applied in stereo reconstruction, but a particle filtering is very flexible probabilistic technique which can easily embed the complex prior models of the reconstructed scene.

The paper is organized as follows: in Section II we introduce the HMM and the likelihood used for the reconstruction, the probabilistic techniques are dealt with in Section III, the examples of the estimated disparity maps are given in Section IV. In Section V, the conclusion and further directions of work are given.

Signals and Systems Group, Faculty of Electrical Engineering, Mathematics and Computer Science, University of Twente, Hogekamp 45, Drienerlolaan 5, 7500 AE Enschede, The Netherlands Tel. +31 53 489 278, Fax.+31 53 489 1060, E-mail: s.damjanovic@ewi.utwente.nl

II. Probabilistic Framework for 3D Reconstruction

We consider the fully calibrated image acquisition setup and where the stereo matching is done along epipolar lines of the rectified images, [3]. We approach the disparity calculation along the epipolar lines as the state estimation problem [1], where the disparity is the discrete state variable. In state estimation, the state variable of interest changes over time. In the application of the state estimation in disparity calculation the state variable evolution happens with the increment of the x-coordinate of the epipolar line in the reference image. Here the time index of the state variable, usually present in the classical state estimation approach, is replaced by the x-coordinate value. Hence, the terms 'previous' and 'next' state are referred to the disparity values with the smaller or larger value of the x-coordinate than the x-coordinate of the observed i.e. 'currrent' disparity denoted by d_x. The length of the state sequence is equal to the number of the pixels in the referent epipolar line for which the disparity values are calculated and it is determined by the size of the window used for the likelihood calculation. If the size of the window is denoted by $W_x = 2 \cdot w_x + 1$ and the length of the referent epipolar line by L, then the x-coordinate takes consecutive values from the array $[w_x + 1,\ L - w_x]$.

Application of the probabilistic algorithms in the state estimation requires a knowledge of the prior and the likelihood. The prior is determined by the chosen HMM. The likelihood is a measure that the state variable is at a certain state and we derive it using NCC coefficients..

The prior knowledge about the state variable, the disparity, is given by the HMM. The number of states of the HMM is equal to the range of disparities i.e. every possible disparity value is represented as a state in the state-space model. The discrete state variable d_x takes values from the set $\Omega_d = \{K_{min}, K_{min}+1, K_{min}+2, ..., K_{max}\}$, where K_{min} and K_{max} are the minimum and the maximum disparities between images. The number of different states is equal to the number of different possible disparity values, $K = K_{max} - K_{min} + 1$.

We consider the common behaviour of the disparity values along epipolar line to properly choose the HMM. The disparity value stays the same for flat frontoparallel surfaces of the scene, and increases or decreases for slanted, along the epipolar line. The increase/decrease of the disparity with reference to the previous disparity value is most often moderate and changes just for several units and rarely large e.g. 10 units or more. We have chosen the shape of transition probabilities as illustrated in Figure 1 to embed this behaviour into the HMM. The variable has the highest probability to stay in the same state. The probability sharply decreases with difference $\Delta = |d_{x+1} - d_x|$ up to $\Delta = trans_{max}$. The probability density for transitions with greater difference $\Delta \in [trans_{max}, jump_{max}]$, i.e. 'jumps', is smaller and constant. The probability density of the outlier transitions, when the state change is greater than $jump_{max}$, is rather small or can be neglected. The total probability of the jump transitions and the total probability of the outlier transitions are denoted as P_{jump} and P_{out}, respectively.

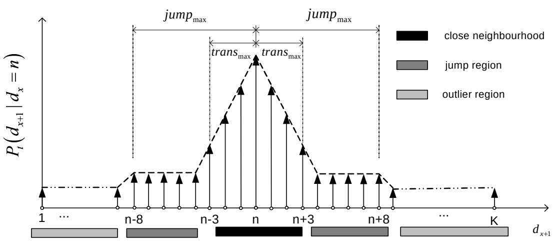


Figure 1. HMM: Transition probability

The likelihood values between epipolar lines $p(z_{1,x}, z_{2,x+d_x} | d_x = k)$, $k \in [K_{min}, K_{max}]$, $x \in [w_x + 1, L - w_x]$ represent the similarity measure of the quadratic $W_x \times W_x$ windows of pixels $z_{1,x}$ surrounding the pixel in the reference image at the position x and windows $z_{2,x+d_x}$surrounding the pixel at the position $x + d_x$ in the other image of the stereo pair of the epipolar line. The NCC is good measure of the similarity between two pixel-windows in images and it inherently compensates for their different offsets and gains [4], [3]. Using the NCC coefficients $NCC(x, d_x = k)$, $k \in [K_{min}, K_{max}]$, $x \in [w_x + 1, L - w_x]$ as likelihood is not suitable, firstly, the NCC coefficients can be negative, while likelihood should be always nonnegative, and secondly, the small ratio between NCC coefficients is not sufficient for good differentiation between different state values. We found that the following simple transformation keeps the same order of the values of the coefficients and gives suitable likelihood values:

$$p(z_{1,x}, z_{2,x+d_x} | d_x = k) \propto \frac{1}{1 - NCC(x, d_x = k)}. \quad (1)$$

III. Probabilistic Stereo Reconstruction algorithms

The stereo reconstruction is defined as the optimal estimation of the disparity state variable along the epipolar line $d_x, x = w_x+1, \ldots, L-w_x$ using the HMM and the likelihood function (1). There are two types of the state estimations: online and offline. If the estimation of the state is done online (in real-time), the estimate is done on the base of the previous and the current measurements as in the case of the forward algorithm and the particle filtering. If the estimation is done offline, it is done not only using the measurements from past and present, but also using the measurements which occurred after the time instant of the state being estimated. The offline estimation is done by back propagation through the measurement sequence as in the backward algorithm and smoothing. The role of the time index is substituted by x-coordinate of the pixel position in the image, see Section II. As all pixels values are known, we can apply both on line and offline algorithms.

Using the forward algorithm, the resulting disparity values are calculated using the maximum a posterior (MAP) criterion. The forward/backward algorithm results in disparities along the epipolar lines whose individual disparities have the minimal error, while the Viterbi algorithm provides the most likely sequence of disparity values.

MAP estimation is also done within the particle filtering framework. The estimation of the posterior probabilities or the filtering distribution can be achieved using the standard filtering recursions via Chapman-Kolmogorov equation

$$p(d_{x+1} | Z(x)) = \int p(d_x | Z(x)) \cdot p_t(d_{x+1} | d_x) d(d_x) \quad (2)$$

and via Bayes rule for the prior update

$$p(d_{x+1} | Z(x+1)) = \frac{p(z_{1,x+1}, z_{2,x+1+d_{x+1}} | d_{x+1}) p(d_{x+1} | Z(x))}{p(z_{1,x}, z_{2,x+d_x} | Z(x))}. \quad (3)$$

In (3) $Z(x) = \{(z_{1,i}, z_{2,i+d_x})_{i \le x, d_x = K_{min}, \ldots, K_{max}}\}$ i.e. $Z(x)$ represents pairs of patches in images for which the disparity is calculated. The approximation strategy for the posterior probability density (2) is the sequential Monte Carlo method, known as particle filter. Within the particle filter framework, the filtering distribution is approximated with an empirical distribution formed of the point of masses, or particles [2], so the posterior probability distribution is given by the particle approximation as

$$p(d_x | Z(i)) \simeq \sum_{n=1}^{N_{cond}} w_{i,norm}^{(n)} \delta(d_x - d_x^{(n)}) \quad (4)$$

where $\delta(\cdot)$ is the Dirac delta function, N_{cond} is number of particles and the normalized importance weights $w_{i,norm}^{(n)}$ are chosen as

$$w_i^{(n)} = p(z(x) | d_x^{(n)}) \quad \text{and} \quad \sum_{n=1}^{N_{cond}} w_{i,norm}^{(n)} = 1, \quad (5)$$

$$w_{i,norm}^{(n)} > 0, \; z(x) = (z_{1,x}, z_{2,x+d_x}).$$

The number of particles N_{cond} should be sufficiently large in order to represent properly the posterior distribution by means of set of samples. For the purpose of comparison with the forward algorithm, the state estimation is done by the MAP criterion. The processing time and number of operation necessary for particle filtering is directly proportional to $I \cdot N_{cond}$.

Smoothing can be performed recursively backward in time using the smoothing formula

$$p(d_x | Z(x)) = \int p(d_{x+1} | Z(x)) \frac{p(d_x | Z(x)) p_t(d_{x+1} | d_x)}{p(d_{x+1} | Z(x))} d(d_{x+1}) \quad (6)$$

Smoothing is applied in addition to particle filtering for generating the realizations of the entire smoothing density $p(d_{x=w_x+1,\ldots,L-w_x}|Z(i))$ based on the forward particle filtering results, [2].

IV. Experiments

Stereo reconstruction using probabilistic techniques is done on the rectified stereo pair shown in Figure 2, [5], [6]. For the stereo pair the ground truth disparity map is known and shown in Figure 3. The disparity values are in the range from 19 to 99, while the occluded pixels are given the value 0, so we choose a HMM with $K_{min} = 19$ and $K_{max} = 99$. We select the parameters for the state transition probabilities: $P_{out} = 0$, $P_{jump} = 0.05$, $jump_{max} = 8$ and $trans_{max} = 3$. The size of the windows for the calculation of the NCC coefficients is $W_x = 31$. The number of particles used in the particle filtering is $N_{cond} = 1000$ and in the smoothing step $N_{back} = 100$.

The disparity map is calculated using dynamic programming and probabilistic techniques. The resulting disparity maps together with the absolute error maps with reference to the ground truth are shown in the figures 4, 5, 6, 7, 8 and 9.

The recovered disparity maps and absolute error maps are very similar. This confirms that the probabilistic techniques can be successfully applied to the stereo reconstruction. The similar errors are the consequence of the same causes. Namely, the size of the windows used in the calculation of the NCC coefficients and correspondingly likelihood values, limits more precise depth calculation. Smaller windows would give better results concerning occlusion detection. Also, the perspective distortion is not taken into account. As well, the simple prior does not model occlusions.

The smooth parts of the scene (the ball and the pins), are reconstructed very well with small error (dark regions in absolute error maps in figures 4b), 5b), 6b), 7b), 8b) and 9b)) and we expect that the probabilistic algorithms can be successfully applied to the stereo reconstruction of the objects with smooth surfaces e.g. reconstruction of faces.

The quantitative quality comparison is given in Table I. For all cases the percentages of the recovered disparity values which are identical to the ground truth disparities are given in column Q_0, while the columns Q_1 and Q_2 show the percentages of the recovered disparities within the range of ± 1 and ± 2 of the ground truth values. The percentage values for particle filtering result are quite comparable with those of other algorithms. The values along the columns are approximately equal. It would be expected that the algorithms which include the explicit or implicit back propagating step: the forward/backward algorithm, the particle filtering followed by smoothing and the Viterbi algorithm, recover the disparity map more accurately. This will be the subject of our further investigations.

V. Conclusion and Further Work

We demonstrated that HMMs and particle filtering can successfully be applied to stereo reconstuction. In these first experiments we used the simple HMM and the likelihood

(a) Left image

(b) Right image

Figure 2. Stereo pair: *Bowling2*, [5], [6]

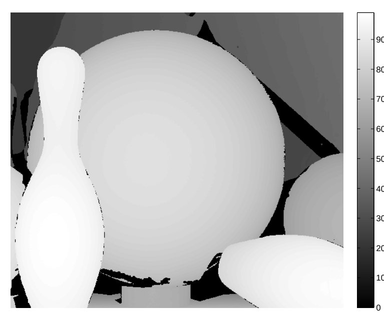
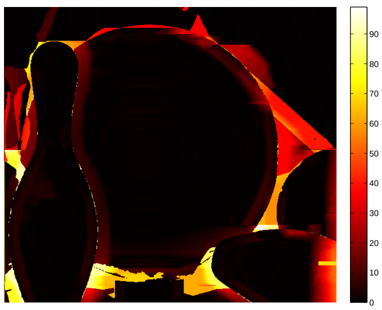


Figure 3. Disparity ground truth, [5], [6]

Table I
Quality of disparity maps

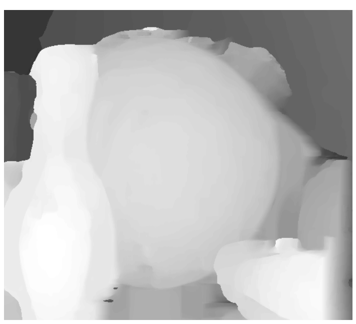
[%]	Q_0	Q_1	Q_2
dynamic programming	43	67	72
forward alg.	47	68	72
forward/backward alg.	46	68	72
Viterbi alg.	48	69	73
particle filtering (PF)	45	68	72
PF with smoothing	43	66	71

(a) Disparity map

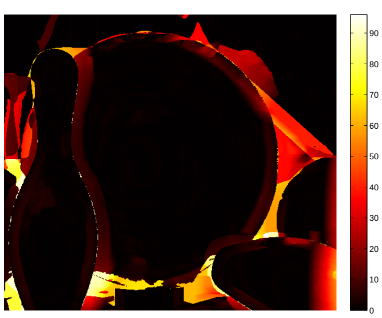


(b) Error map

Figure 4. Dynamic programming result



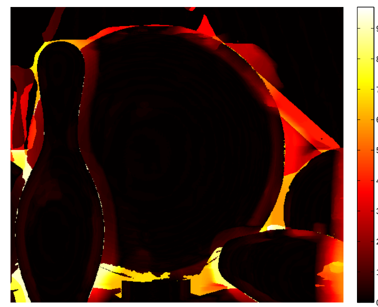
(a) Disparity map



(b) Error map

Figure 5. Forward algorithm result

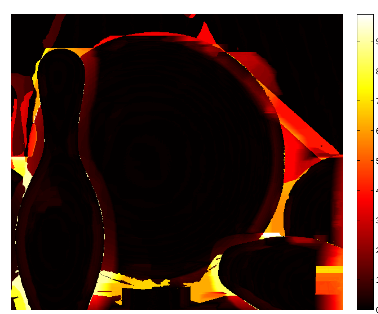
(a) Disparity map



(b) Error map

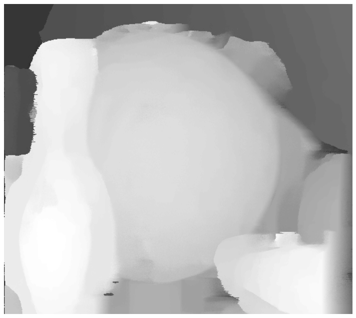
Figure 6. Forward-backward result

(a) Disparity map

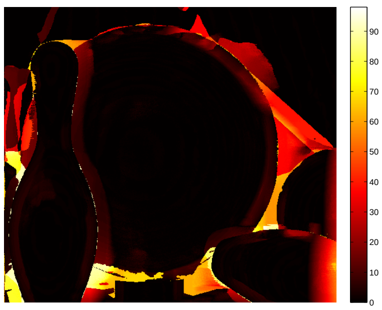


(b) Error map

Figure 7. Viterbi algorithm result



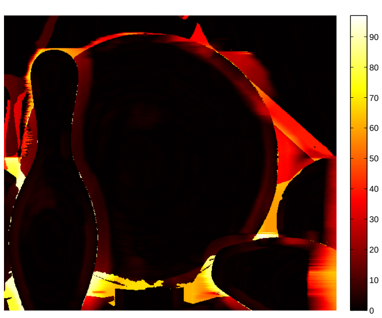
(a) Disparity map



(b) Error map

Figure 8. Particle filtering result

(a) Disparity map



(b) Error map

Figure 9. Smoothing result

based on the NCC coefficients. The performance of the particle filtering is comparable to the performance of dynamic programming.

We plan to further improve the performance of this probabilistic 3D reconstruction technique by including more complex prior and more advanced likelihood. We expect a significant improvement of the quality of reconstruction if we exploit the flexibility of the particle filtering with respect to including only scene knowledge.

REFERENCES

[1] F. van der Heijden, R. Duin, D. de Ridder, and D. Tax, *Classification, Parameter Estimation and State Estimation An Engineering Approach Using MATLAB*. WILEY, 2004.

[2] A. D. W. Fong, S.J. Godsill and M. West, "Monte Carlo smoothing with application to audio signal enhancement," *IEEE Transaction on Signal Processing*, vol. 50, pp. 438–449, February 2002.

[3] O. Faugeras, *Three-Dimensional Computer Vision - A Geometric Viewpoint*. The MIT Press, 1993.

[4] M. Z. Brown, D. Burschka, and G. D. Hager, "Advances in Computational Stereo," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, no. 8, pp. 993–1008, 2003.

[5] H. Hirschmüller and D. Scharstein, "Evaluation of cost functions for stereo matching," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2007)*, 2007.

[6] "http://vision.middlebury.edu/stereo/data/."