Temporal Filtering of Disparity Measurements

L. Di Stefano, S. Mattoccia, G. Neri, D. Piccinini
DEIS, University of Bologna
Via Risorgimento 2, 40136 Bologna, Italy
{ldistefano, smattoccia, gneri}@deis.unibo.it
dpiccinini@thesis.deis.unibo.it

Abstract

The paper proposes a temporal filtering technique for the disparity measurements generated by area-based stereo-matching algorithms. The technique improves temporal consistency of disparity measurements by reducing the matching errors due to the noise affecting the imaging system. Moreover, the technique is capable of increasing the number of correct matches by locating uncertain measurements with a criterium based on statistical assumptions that has proven to be more accurate and selective than those relying on texture operators only which are typically deployed with standard area-based stereo algorithms.

1. Introduction

With the advent of PC-based real-time stereo matching [11, 6, 14] reliability of range data becomes an issue since an ever increasing number of applications rely on stereo to infer $3D$ information on the surroundings. Hence, a significant problem that arises out of processing stereo sequences with standard area-based stereo algorithms [11, 6] is the noise-generated, spurious temporal variation of disparity measurements. This problem can be easily observed in the case of a static scene (i.e. a scene without moving objects) viewed by a fixed stereo-imaging system: ideally the disparity measurements at each image point should be constant over time while, due to the noise affecting the imaging system, they vary significantly. This effect has been pointed out and characterized by Matthies and Grandjean [13]. The interested reader can found an example of such a behaviour at the web page [2] (disparity measurements are encoded with brighter grey levels for closer points and darker for farther points; unmatched points are represented in red).

This paper proposes a filtering technique - referred to as *Temporal Filter* - capable of reducing the spurious temporal variations of disparity measurements obtained by area-based stereo algorithms by constraining such measurements to be coherent over time. Moreover, the technique can increase the number of correct matches and, though our experimental data have been collected using two specific stereo algorithms, namely SVS [11] and VSA [6], it is suitable to every stereo matching algorithm yielding dense disparity maps.

We assume that cameras are in a fixed position with both static and moving objects in the field of view. Examples of this configuration can be found in stereo based video surveillance systems [4, 5, 10], person counting systems [3] and many others real applications.

2. Dealing with Temporal Variations by Texture Operators

Standard real-time, area-based stereo algorithms, like those considered in this paper, detect and discard uniform regions using operators that evaluate the degree of texture at each image point. These *texture operators* discriminate between textured and low textured areas by means of a threshold value. SVS uses an *interest operator* [11] while VSA uses the variance of the intensity values in a neighbourhood of each pixel [6, 7].

Since most spurious temporal variations of disparity measurements are found in low-textured regions, the typical approach to dealing with disparity variations when using such algorithms consists in acting on the threshold of the texture operator, which is chosen to be as selective as to discard most of the image points that present excessive temporal variation of disparity.

However, increasing the selectivity of the texture operator to reduce temporal variations has the side-effect of eliminating also many points that could be matched correctly, since there is not a direct relationship between the matching process and the logic of the texture operator. In addition, temporal variation of disparities do not appear only in low textured regions but in any region that is sensitive to noise

(e.g. in regions with repetitive patterns). This is the major drawback of using *texture operators* to deal with temporal variations, since points are filtered away with *a priori* assumptions on their temporal behaviour based on the degree of texture. This is a conservative choice since allows for discarding many potentially ambiguous points during the matching phase but at the same time discards many potentially good matches assuming that low-textured points are necessarily prone to mismatches. This assumption, though frequently verified, it is not always true since other parameters like *distinctiveness* [12] should be taken into account in order to obtain more reliable confidence measurements. It is worth pointing out that in general ambiguous points to match are those producing disparity maps more sensitive to noise perturbation. Nevertheless texture operators attempt to extract confidence information without taking into account any feedback resulting from the disparity maps sequences generated by the matching algorithm.

The proposed technique deploys the *behaviour* of disparity measurements over successive frames in order to selectively eliminate those matches that presents high probability to be wrong. Decisions on the reliability of disparities are taken comparing the temporal behaviour of the disparity maps with a reference model built upon statistical observations. Hence, the proposed temporal filtering technique is capable of increasing the number of correct matches since it allows for keeping the selectivity threshold of the texture operator of the matching algorithm at much lower values if compared with those that should be used to reduce the number of unstable disparity measurements.

3. Classification of Measurement Errors

In this Section we present a classification of the major errors affecting the disparity maps generated by area-based stereo matching algorithms in the case of a static scene. This classification will be exploited to define the structure of our temporal filtering technique.

As already discussed, the disparity measurements computed by area-based stereo algorithms are affected by a significant number of errors as well as by spurious variations over successive frames. These wrong measurements can be classified into two categories based of their temporal behaviour: *systematic* and *random errors*.

Systematic errors occurs when the disparity measurement is wrong but stationary over successive frames. This implies that these kind of errors are not affected by the noise introduced by the imaging system. Systematic errors are due to several causes such as misalignment, foreshortening [15], ambiguous matching patterns, *object growing effects* [9]. Some of these errors can be compensated adopting specific techniques. However, the common element of these errors relies in their independence from image noise. Our filtering technique is not designed to deal with these kind of errors and therefore they won't be considered in the remainder of this paper

Conversely, *random errors* occurs when the depth measurements produce non stationary values over frames. These are mostly due to the image noise that result in a perturbation of the matching process. Thus, disparity values associated with the same point in two consecutive frames of a static scene are different resulting in a fluctuation of the disparity map sequence that can be easily verified by visual observation. The same behaviour can be found in static areas of dynamic scenes. The impact of noise on the quality of range data have been studied by Matthies and Grandjean [13]. Another source of uncertainty of the disparity estimates is the correlation between errors for neighbouring pixels [13]. It is worth point out that typical noise fluctuations between frames are generally greater than those between neighbourhood pixels of the same frame.

We have observed that points of the disparity map affected by random errors can be further divided on the basis of their temporal behavior in the following two classes:

a) Points showing *spurious variation of disparity*; these points, though constantly matched by the algorithm, produce spurious variations of disparity over successive frames. It has been shown that in this case the disparity value at a given point behaves like a random variable characterized by a Gaussian distribution [13].

b) Points which are *alternatively matched or unmatched* by the algorithm. This behavior occurs for example near the occlusions associated with depth discontinuities. Moreover, noise randomly modify the degree of texture at a given point so that in low textured regions the output of the matching algorithm shows an alternation over successive frames between unmatched and - typically wrong - matched disparity values. We have observed that sometimes the left-right consistency check [8], which is a common matching constraint used in area-based stereo, can discard these points but also that very frequently this constraint is not effective, resulting in the typical flickering shown by numerous points in real-time sequences of disparity measurements.

4. Temporal Filtering for Static Scenes

The previous classification allow us to define a set of rules aimed at reducing temporal variations of disparity measurements in the case of static scenes. The basic idea is that by analyzing the temporal evolution of disparity values for a certain number of consecutive frames it is possible to infer information about reliability of matches. Since we are interested in keeping trace of the temporal statistic associated with each single match in order to assess its reliability, we need to define a framework with significant parameters. In the previous section we have characterized random er-

rors describing the two typical behavior with static scenes: points showing *spurious variation of disparity* and points which are *alternatively matched or unmatched*.

We denote with Φ the number of consecutive disparity maps used to keep history of the behavior of each points. In the remainder the number of frames used to analyze disparity maps evolution will be referred to as *order of the filter*. The number of commutations between matched and unmatched disparity values at a given point is expressed as $\Gamma_t(i,j)$. If a point has shown more than Γ_{max} commutations within the last Φ frames its disparity value is set to unmatched in the current frame. Parameter $\Omega_t(i,j)$ represents the number of frames within the last Φ in which the point has been matched. Thus, one point is considered as reliable if it shows a low $\Gamma_t(i,j)$ and an high $\Omega_t(i,j)$. Conversely, if the point is matched less than Ω_{min} times is discarded as unreliable by the filter and marked as unmatched. The third parameter of the filter is $\Delta d_t(i,j)$, which is used to evaluate the temporal variation of disparity $d_t(i,j)$ at a given point (and, as usual, within the temporal window of length Φ). It represents the average sum of the absolute value of the differences between consecutive matched values:

$$\Delta d_t(i,j) = \frac{\sum_{k=0}^{\Phi-1} |d_{t-k}(i,j) - d_{t-k-1}(i,j)|}{\phi} \quad (1)$$

In the sum of equation (1) if one or both terms are undefined, since disparity has been left unmatched by the stereo algorithm, its contribution is not taken in account (i.e. $|d_{t-k}(i,j) - d_{t-k-1}(i,j)| = 0$). The parameter $\Delta d_t(i,j)$ give an indication on how smoothly the disparity measurements at a given point vary over time. If one point has $\Delta d_t(i,j)$ greater than Δd_{max} it is considered unreliable by the filter. The logic of the filter can be summarized as follow: one disparity value is considered reliable if the following three conditions are verified:

$$d_t(i,j) \quad reliable \quad if \begin{cases} \Gamma_t(i,j) \leq \Gamma_{max} \\ \Omega_t(i,j) \geq \Omega_{min} \\ \Delta d_t(i,j) \leq \Delta d_{max} \end{cases} \quad (2)$$

since in the last Φ frames the behavior of the disparity values for the point examined were those expected for a correct match. Conversely, if the point does not pass one of the three tests the disparity value generated by the algorithm is discarded since in the last Φ frames the disparity values for that point differ from those expected for a correct match. Moreover, if one point is left unmatched in the current frame but in the previous frames its values were rather stable, the disparity is set to its previous value. The order of the filter, Φ, affects its capability to take correct decisions. Increasing the order of the filter produce better decision capability but this effect is not perceivable when the order of the filter gets too high. Our experimental results show that orders of about 20 are enough to obtain a correct behavior of the filter.

5. Temporal Filtering for Dynamic Scenes

In this section we extend the temporal filtering technique described so far so as to deal with the more general case of scenes containing moving objects. In presence of moving objects the constraints defined for the case of static scenes are no longer effective since the statistics for each pixel are collected assuming that a pixel corresponds always to the same scene point. But when an object is moving in the scene we can't make any statistical assumption concerning the temporal evolution of disparity measurements and, instead, we should let disparity values by-pass the filter. In such a case the degree of reliability may be kept high, at the cost of a larger amount of potentially good matches discarded, by increasing the selectivity threshold of the texture operator. On the other hand, as pointed out in the previous section, the filtering capability within the static areas of the scenes provided by the temporal filter gets better when the order of the filter increase. Thus, rather than using a trade-off between these two opposite requirements, we use a simple motion detection operator to find out whether a point in the current frame is a moving point or a static point and then adjust dynamically the order of the filter according to a state information associated with each image point.

Thus, based on the state of each point in the current frame, the temporal filter for dynamic scenes uses a high-order temporal constraint when the point is static and avoids any constraint (i.e. uses an order of 0) when the point is moving. The temporal filter requires two values for the threshold of the texture operator: a low one (poorly selective) for static points and an high one (highly selective) for moving points, so as to increase the degree of reliability when the filter is by-passed. The disparity map is computed by the stereo algorithm using a low texture threshold and then analyzed by the temporal filter considering the state of each point. Moving points with texture under the higher threshold are discarded by the temporal filter.

It is worth pointing out that the two transition of state $Static \rightarrow Dynamic$ and $Dynamic \rightarrow Static$ are completely different. Transition $Static \rightarrow Dynamic$ implies that the order of the filter must drop to zero as soon as motion is detected in order to get rid of the past history of the point under examination. As for transition $Dynamic \rightarrow Static$, the order of the filter is incremented of one unit at each frame, as long as no motion is being detected, in order to collect correct information on the behavior of a point that was previously moving.

While a point is in a state belonging to the $Dynamic \rightarrow Static$ transition most of the parameters used by the filter (i.e. texture threshold, Γ_{max}, Ω_{min}) are scaled according to the order associated with the state.

If a point remains in a static state for Φ frames the order of the filter reaches its maximum value and the filter can

judge the temporal behavior of the point on the basis of the maximum amount of information concerning past frames. While the order is growing towards the maximum value the filter improves its decisions as the order gets higher.

Hence, in the dynamic version the order of the filter is dynamically changed at each point and depends on the previous state and on the motion information provided by the motion detection operator. The state diagram of Figure 1 describes graphically the behaviour of the temporal filter in case of dynamic scenes.

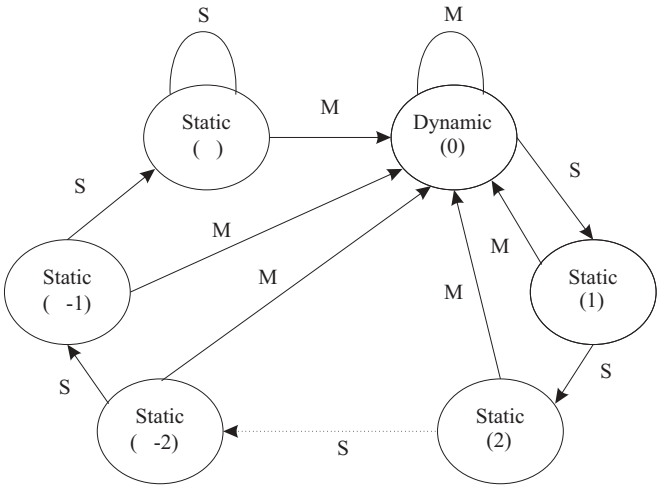

Figure 1. State diagram of the temporal filter.

The diagram is relative to a generic point, with the state transitions driven by the output of the motion detection operator at the point: M means that motion has been detected while S means static point (i.e. motion has not been detected). The diagram shows also, within each state, the motion condition (Dynamic or Static) and the order of the filter. Motion detection is carried out using a simple operator based on thresholding the absolute difference between two consecutive frames of the left-view of the stereo pair. This allows for retrieving motion information at a very low computatational cost. The temporal filter has been implemented by means of a very efficient, incremental computational scheme which is suitable to real-time gathering of disparity measurements.

6. Experimental Results

The temporal filtering technique described in this paper has been tested with various real stereo sequences showing its ability to yield a significant reduction of disparity fluctuations and as well as to increase the number of points matched correctly.

In the case of the static sequence of Figure 2 (available at the web site [2]) the technique reduces the average temporal

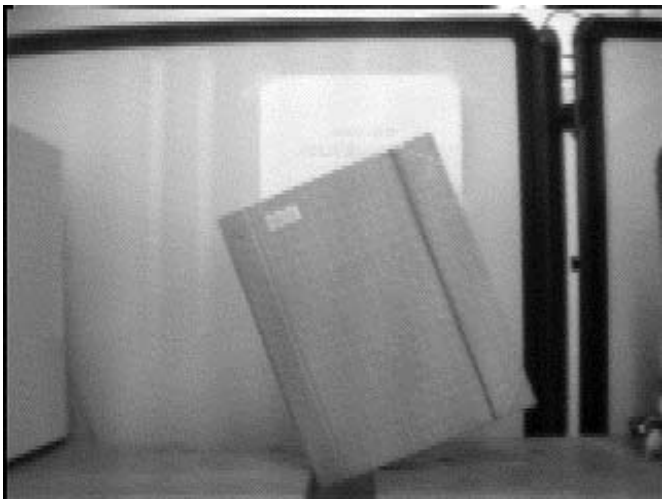

Figure 2. One frame of the static sequence (Left Image).

variance of disparity measurements from 75.1 to 18.3 and increases the number of correct matches with the SVS algorithm of more than 10%. The significantly better stability of disparity measurements provided by the filter can be perceived visually looking at the videos available at [2], which show the disparity map sequences obtained with SVS and with the filter respectively disabled and enabled. Figure 3 and Figure 4 show the disparity maps yield by SVS on one frame of the sequence of Figure 2. The larger number of matched points in the background and foreground regions are clearly visible (unmatched points are represented in red, which appears as the darkest gray if this paper is printed in b/w).

In the remainder of this Section we show the experimental results obtained on a simple sequence consisting of a parallelepipidal object translating in a direction parallel both to the camera's image plane and to a flat background. The sequence is composed of 500 frames: in the first 100 frames the scene is static, in the following 300 frames the object is moving while in the last 100 frames the scene again is static. Table 1 reports the results obtained with the two stereo algorithms considered in this paper (SVS and VSA) without the temporal filter enabled and disabled. The radius of the correlation windows used in the matching step was 6 pixels and the disparity range of 32 pixels.

The first two rows of Table 1 allows for comparing the results obtained running the standard SVS algorithm with those obtained using the algorithm in conjunction with our temporal filter. The parameters for the temporal filter were $\Phi = 20$, $\Gamma_{max} = 6$, $\Omega_{min} = 17$, $\Delta d_{max} = 4$. The setting of the texture operator of SVS, when not used in conjunction with the temporal filter, was obtained following the

Alg.	Filter	Texture Oper.	Correct Matches	Wrong Matches	Not Matched
SVS	Off	12	38.40%	0.35%	61.25%
SVS	On	2-10	44.62%	0.30%	55.08%
VSA	Off	6	25.45%	0.51%	74.04%
VSA	On	2-6	29.67%	0.37%	69.96%

Table 1. Dynamic scene: experimental results.

Figure 3. Disparity map of Figure 2 with temporal filter disabled.

Figure 4. Disparity map of Figure 2 with temporal filter enabled.

usual criterion of observing the disparity maps generated in real-time and minimising the number of points which present excessive temporal variation. The value found according to this criterion was 12. With the temporal filter enabled, second row, maximum and minimum values for the texture operator were set respectively to 2 for static points and 10 for moving points. Table 1 shows the improvements associated with the use of the proposed filter: the number of correct matches increases from 38.40% to 44.62% while the number of wrong matches remains almost stable (decreasing from 0.35% to 0.30%). It is worth pointing out that most of the new points matched when using the filter belong to the background, since the background is static and the filter allows for discarding more selectively wrong disparity measurements in static regions. The original sequence and the disparity map sequences obtained using SVS with the filter enabled and disabled can be viewed at the web site [1]. The last two rows of Table 1 provide similar results in the case of the VSA algorithm. The texture operator used by VSA is the gray-level variance, and the best setting for the considered sequence when using the algorithm without the temporal filter was found to be 6; with filter enabled the two variance threshold were set to 2 and 6. Also with VSA the use of the filter improves disparity measurements since it allows for increasing the number of correct matches from 25.45% to 29.67% and for decreasing the number of wrong matches from 0.51% to 0.37%.

7. Conclusion

In this paper we have addressed the problem of spurious temporal fluctuation of disparity measurements due to the image noise affecting any stereo imaging system. This is an important issue in real-time stereo vision systems available nowadays on standard personal computers. We have classified the behavior of spurious fluctuations by analyzing their temporal evolution and starting from the simpler case of scenes containing only static objects. This preliminary study allowed us to define a set of rules and a framework aimed at the selection of unreliable matches based on their behaviour over successive frames. The general case of scene containing both static and dynamic objects has been addressed deploying motion detection information and allowing the filter to change dynamically its order. That is, for

moving objects the filter is by-passed (i.e. the order of the filter is set to 0) so as to promptly follow the structural disparity variations in the scene. Conversely, for static points the order of the filter is kept as high as possible, through a smooth incremental variation, in order to provide the maximum capability to filter away uncertain disparity measurements. To assess the performance of the proposed filtering technique we have considered two area-based stereo algorithms, namely SVS and VSA. Our experimental results show that the filter is capable of reducing the spurious temporal variations of disparity measurements as well as of increasing the number of points matched correctly by localising uncertain matches more selectively with respect to the use of texture operators provided by standard area-based stereo algorithms.

References

[1] Experimental results: dynamic scenes. http://labvisione.deis.unibo.it/~smattoccia/Temporal/TemporalDynamic.html.

[2] Experimental results: static scenes. http://labvisione.deis.unibo.it/~smattoccia/Temporal/TemporalStatic.html.

[3] D. Beymer. Person counting using stereo. In *Workshop on Human Motion*, pages 127–133, 2000.

[4] D. Beymer and K. Konolige. Real-time tracking of multiple people using continuous detection. In *IEEE Frame Rate Workshop*, Corfu, Greece, 1999.

[5] T. Darrell, M. Gordon, M. Harville, and J. Woodfill. Integrated person tracking using stereo, color and pattern detection. In *IEEE Conf. on Computer Vision and Pattern Recognition*, pages 601–608, 1998.

[6] L. Di Stefano and S. Mattoccia. Fast stereo matching for the videt system using a general purpose processor with multimedia extensions. In *Int. Workshop on Computer Architecture for Machine Perception*, Padova, Italy, Sept. 2000.

[7] O. Faugeras et al. Real-time correlation-based stereo: algorithm, implementation and applications. INRIA Technical Report n. 2013, 1993.

[8] P. Fua. Combining stereo and monocular information to compute dense depth maps that preserve depth discontinuities. In *12th. International Joint Conference on Artificial Intelligence*, pages 1292–1298, Sydney, Aug. 1991.

[9] S. Gautama, S. Lacroix, and M. Devy. Evaluation of stereo matching algorithms for occupant detection. In *RATFG99*, pages 177–184, 1999.

[10] I. Haritaoglu, D. Harwood, and L. Davis. W4s: A real time system for detecting and tracking people in 2.5 d. In *Fifth European Conference on Computer Vision*, Freiburg, Jun 1998.

[11] K. Konolige. Small vision systems: Hardware and implementation. In *8th Int. Symposium on Robotics Research*, Hayama, Japan, 1997.

[12] R. Manduchi and C. Tomasi. Distinctiveness maps for image matching. In *Proceedings of 10th International Conference on Image Analysis and Processing*, Venice, Italy, Sept. 1999.

[13] L. Matthies and P. Grandjean. Stochastic performance modeling and evaluation of obstacle detectability with imaging range sensors. *IEEE Trans. On Robotics and Automation*, 10:783–792, Dec 1994.

[14] Point Grey Research. Digiclops stereo vision system. http://www.ptgrey.com, 2000.

[15] Y. Xiong and L. Matthies. Error analysis of a real-time stereo system. In *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 1997.