Permanency Memories in Scene Depth Analysis

Miguel A. Fernández[1], José M. López-Valles[2], Antonio Fernández-Caballero[1],
María T. López[1], José Mira[3], and Ana E. Delgado[3]

[1] Universidad de Castilla-La Mancha
Escuela Politécnica Superior de Albacete, 02071 - Albacete, Spain
{miki,caballer,mlopez}@info-ab.uclm.es
[2] Universidad de Castilla-La Mancha
Escuela Universitaria Politécnica de Cuenca, 13071 - Cuenca, Spain
JoseMaria.Lopez@uclm.es
[3] Universidad Nacional de Educación a Distancia
E.T.S.I. Informática, 28040 - Madrid, Spain
{jmira,adelgado}@dia.uned.es

Abstract. There are several strategies of how to retrieve depth information from a sequence of images, like depth from motion, depth from shading and depth from stereopsis. In this paper, we introduce a new method to retrieve depth based on motion and stereopsis. A motion detection representation helps establishing further correspondences between different motion information. This representation bases in the permanency memories mechanism, where jumps of pixels between grey level bands are computed in a matrix of charge accumulators. For each frame of a video stereovision sequence, the method fixes the right permanency stereo memory, and displaces the left permanency stereo memory by pixel on the epipolar restriction basis over the right one, in order to analyze the disparities of the motion trails calculated. By means of this functionality, for all possible displacements of one permanency memory over the other, the correspondences between motion trails are checked, and the disparities are assigned, providing a way to analyze the depths of the objects present in the scene.

1 Stereovision-based Depth Analysis

In general, there are several strategies of how to retrieve depth information from a sequence of images, like depth from motion, depth from shading and depth from stereopsis. In this paper, we introduce a new method to retrieve depth based on motion and stereopsis. In a conventional stereoscopic approach, usually two cameras are mounted with a horizontal distance between them. Consequently, objects displaced in depth from the fixation point are projected onto image regions, which are shifted with respect to the image center. The horizontal component of this displacement can be used to determine the depth of the object. Due to the geometry of the optic system, and considering the epipolar constraint, it is thereby sufficient to restrict disparity analysis to the projection of corresponding linear segments in the left and right camera. In some approaches, the disparity

is computed by searching the maximum of the cross correlation between image windows along the epipolar lines of the left and right image [1]. Similarly, this can be done by trying to match discernible image features.

So far, many algorithms have been developed to analyze the depth in a scene. Brown et al. [2] describe a good approximation to all of them in their survey article. In many previous works, a series of restrictions are used to approach the correspondence problem. The most usual restriction is the disparity restriction, which considers that is not probable that there exist objects very close to the camera. The scene uses to be limited to a medium distance. This way, too high disparities are eliminated [3]. Koenderink and van Doorn [4] expressed the necessary theory in best initial works related to disparity restriction, and Wildes [5] implemented some of their ideas [6]. More recently, disparity in stereoscopy continues showing its great interest (e.g., [7], [8]).

All these developments approach the depth analysis by different methods; but most of them have as a common denominator that they work with static images and not with motion information. In this paper, we have chosen as an alternative not to use direct information from the image, but rather the one derived from motion analysis. The system proposed uses as input the motion information of the objects present in the stereo scene, and uses this information to perform a depth analysis of the scene.

2 Motion Detection from Permanency Memories

The input to our system is a pair of stereo image sequences. The sequences have been acquired by means of two cameras arranged in a parallel configuration. The central idea behind this approach is to transpose the spatially defined problem of disparity estimation into the temporal domain and to compute the disparity simultaneously with the incoming data flow. This can be achieved realizing that in a well-calibrated fronto-parallel camera arrangement the epipolar lines are horizontal and thereby identical to the camera scan-lines. Our team has already tested the motion analysis algorithm used in this work in monocular video sequences ([9],[10],[11]).

In this case, motion analysis performs separately on both stereovision sequences in two phases. The first analysis phase is based in grouping neighboring pixels that have similar grey levels in closed and connected regions in an image frame. The method used is segmentation in grey level bands. This method consists in reducing the resolution of illumination levels of the image, obtaining this way a lower number of image regions, which potentially belong to a single object in motion. The second phase has to detect possible motions of the segmented regions through the variation of the grey level band of the pixels.

After motion detection, we now introduce a representation that may help to establish further correspondences between different motion information. This representation finds its basis in the permanency effect. This effect considers the jumps of pixels between bands, and it consists of a matrix of charge accumulators. The matrix is composed of as many units in horizontal and vertical direction as

pixels there are in an image frame. Initially all accumulators are empty; that is to say, their charge is zero. When a jump between grey level bands occurs at a pixel, the charge unit (accumulator) of the permanency memory at the pixel's position is completely charged. After the complete charge, each unit of the permanency memory goes decrementing with time (in a frame-by-frame basis) down to reaching the minimum charge value, while no motion is detected, or it is completely recharged, if motion is detected again. Fig. 1 shows all these issues. Fig. 1a and Fig. 1b show two images of a monocular sequence. The advance of a car may be appreciated, as well as a more slight movement of a pedestrian. In Fig. 1c you may observe the effect of these moving objects on the permanence memory.

Fig. 1. Permanency effect: (a) one image of a sequence, (b) same perspective after some seconds, (c) motion trails as represented on the permanence memory

The difference between a quick object as the car, which is leaving a very long motion trail (from dark grey to white), and a pedestrian whose velocity is clearly slower and whose motion trail is nearly unappreciable with respect to the cars one, is presented. Thus, permanency memories enable representing the motion history of the frames that form the image sequence, that is to say, there is segmentation from the motion of the objects present in the scene.

3   Disparity Analysis from Permanency Memories

Now, motion-based segmentation facilitates the correspondence analysis. Indeed, motion trails obtained through the permanency memories charge units are used to analyze the disparity between the objects in the stereo pair. The set of all disparities between two images of a stereo pair is denominated the disparity map. The key idea is that a moving object causes two identical trails to appear in epipolar lines of the permanency stereo memories. The only difference relies in their relative positions, affected by the disparity of the object at each moment.

Looking at Fig. 2 it is possible to analyze the motion of each one of the three objects present in the permanency memories from their motion trails. This initial analysis is independent of the epipolar constraint studied. You may observe that object "a", which has a long trail and has his maximum charge towards the left, is advancing to the left at a high speed. Object "b", with a shorter trail, is also advancing towards the same direction but at a slower velocity. Finally, object

"c", whose trail is inverted in horizontal, is moving to the right at a medium velocity, as shown by its trail.

Also from Fig. 2, but now comparing between the motion trails in both epipolar lines, disparity is analyzed. Motion trail of object "b" presents a null disparity. Therefore, we can conclude that this trail corresponds to an object that is far away from the cameras. Remember that due to our parallel cameras configuration, pixels with a null disparity are located in the infinite. Object "a" has a little greater disparity. Finally, object "c" offers the greatest disparity.
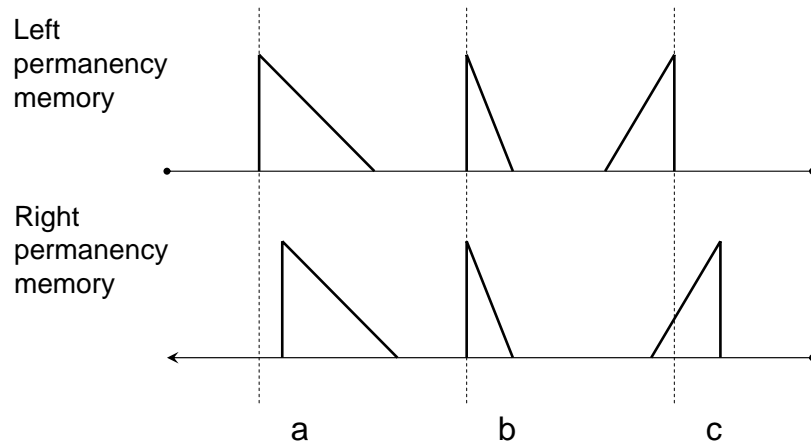
Fig. 2. Disparity of permanency memories

The generalization to global analysis on complete stereo images consists in totally superimposing the two permanency stereo memories under study, and not only their epipolar lines. One of the memories will be displaced over the other looking for motion trails that coincide in both x and y directions. Once the displacement where the coincidence of pixels of motion trails is maximum in size has been calculated, this value is assigned the disparity value. By means of this functionality, for all possible displacements of one permanency memory over the other, the correspondences between motion trails are checked and the disparities are assigned.

4 Data and Results

In order to test our algorithms, the scenario called "IndoorZoom" downloadad at labvisione.deis.unibo.it/ smattoccia/stereo.htm has been used. Fig. 3 shows the result some of the more representative results of applying our algorithms to the "IndoorZoom" scenario. In column (a) some input images of the right camera are shown, in column (b) the segmentation in grey level bands may be appreciated, in column (c) motion information as represented in the right permanency memory

is offered, and in column (d) the final output, that is to say, the scene depth as detected by the cameras, is presented. You may observe on Fig. 3 that clearer
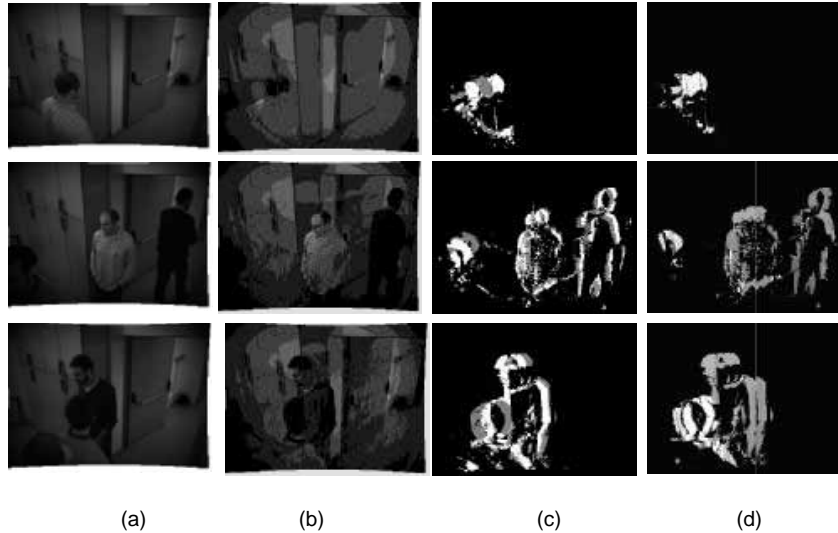

(a)   (b)   (c)   (d)

Fig. 3. Results for "IndoorZoom" scenario

colors means that persons are closer to the cameras. Black means there is no motion detected. Consider the case of occlusions, where, nevertheless, the motion trails, hence, the depths, are different, enabling this way to distinguish among different persons.

5   Conclusions

In this paper, we have introduced a new method to retrieve depth based on motion and stereopsis. A motion detection representation helps establishing further correspondences between different motion information. This representation bases in the permanency memories mechanism, where jumps of pixels between grey level bands are computed in a matrix of charge accumulators. Thus, for the purpose to analyze scene depth from stereo images, we have chosen the alternative not to use direct information from the image, but rather the one derived from motion analysis. This alternative provides an important advantage.

Trough motion information it is easier to use correspondences than by grey level information of the frames. The results are also more accurate and robust. This is due to the instantaneous motion features, such as position, velocity, acceleration and direction of the diverse moving objects. Motion information of an object is different from any other moving object's one. Nonetheless, when observing motion features of a concrete object in both stereo sequences at the

same time instant, we appreciate that these features are extremely similar. This is the reason why it is easy and robust to establish correspondences between the motion information of an object at the right image respect to the object at the left image. There exist very few ambiguity possibilities.

6 Acknowledgements

References

1. Haralick, R.M., Shapiro, L.G.: Computer and Robot Vision. Addison-Wesley (1992)
2. Brown, M. Z., Burschka, D., Hager, G. D.: Advances in computational stereo. IEEE Transactions on Pattern Analysis and Machine Intelligence **25**:8 (2003)
3. Sumi, Y., Kawai, Y., Yoshimi, T., Tomita, F.: 3D object recognition in cluttered environments by segment-based stereo vision. International Journal of Computer Vision **46**:1 (2002) 5–23
4. Koenderink, J.A., van Doorn, A.J.: Geometry of binocular vision and a model for stereopsis. Biological Cibernetics **21** (1976) 29–35
5. Wildes, R.P.: Direct recovery of three-dimensional scene geometry from binocular stereo disparity. IEEE Transactions on Pattern Analisis and Machine Intelligence **13**:8 (1991) 761–774
6. Wilson, R., Knutsson, H.: A multiresolution stereopsis algorithm based on Gabor representation. Proceedings of the IEE International Conference on Image Processing and Applications (1989) 19–22
7. Mhlmann, K., Maier, D., Hesser, J., Mnner, R.: Calculating dense disparity maps from color stereo images, an efficient implementation. International Journal of Computer Vision **47**:1-3 (2002) 79–88
8. Gutirrez, S., Marroqun, J.L.: Robust approach for disparity estimation in stereo vision. Image and Vision Computing **22**:3 (2004) 183–195
9. Fernández-Caballero, A., Fernández, M.A., Mira, J., Delgado, A.E.: Spatiotemporal shape building from image sequences using lateral interaction in accumulative computation. Pattern Recognition **36**:5 (2003) 1131–1142
10. Fernández, M.A., Fernández-Caballero, A., López, M.T., Mira, J.: Lenght-speed ratio (LSR) as a characteristic for moving elements real-time classification. Real-Time Imaging **9** (2003) 49–59
11. Fernández-Caballero, A., Mira, J., Delgado, A.E., Fernández, M.A.: Lateral interaction in accumulative computation: A model for motion detection. Neurocomputing **50** (2003) 341–364