A People Counting System Based on Dense and Close Stereovision

Tarek Yahiaoui[1], Cyril Meurie[1], Louahdi Khoudour[1], and François Cabestaing[2]

[1] French National Institute for Transport and Safety Research (INRETS-LEOST)
20 rue Elisee Reclus, BP 317, F-59666 Villeneuve d'Ascq Cedex
{tarek.yahaoui, cyril.meurie, louahdi.khoudour}@inrets.fr
[2] University of Sciences and Technology of Lille, (LAGIS laboratory, UMR 8146)
Cite Scientifique, F-59655 Villeneuve d'Ascq Cedex
fcab@ieee.org

Abstract. We present in this paper a system for passengers counting in buses based on stereovision. The objective of this work is to provide a precise counting system well adapted to buses environment. The processing chain corresponding to this counting system involves several blocks dedicated to the detection, segmentation, tracking and counting. From original stereoscopic images, the system operates primarily on the information contained in disparity maps previously calculated with a novel algorithm. We show that one can obtain a counting accuracy of 99% on a large data set including specific scenarios played in laboratory and on some video sequences shot in a bus during exploitation period.

1 Introduction

Passengers counting is a very important need for transport operators. Indeed, they need reliable and precise counting information in order to plan and manage in the most appropriate way human, financial and material resources. In particular, the sharing of incomes between operators exploiting the same lines and the fraud rate evaluation are two objectives that require very precise counting information. In this case, the current approximations and statistics are not sufficient to achieve those goals. There are a lot systems for counting people in transit based on various technologies namely: infrared sensors, ultrasound, carpet contact, light rays. However, existing systems achieve too high error rates and can not correctly deal with complex situations like crowded ones. To face this problem, we have developed a counting system based on artificial vision and image processing and composed of several processing blocks exploiting stereovision. In this paper, we present the different components constituting the processing chain of the counting system. We also present the evaluation results of each block of the chain as well as the counting results on a dedicated database.

A. Elmoataz et al. (Eds.): ICISP 2008, LNCS 5099, pp. 59–66, 2008.

2 Proposed Counting System

Counting accurately passengers in a bus is particularly delicate. The use of a single camera system is not sufficient for reasons such as: occlusions, illumination changes and difficult configurations caused by crowd behavior in bus. Hence, we opted for an approach based on dense stereovision [1], [2]. This approach is less sensitive to illumination changes and could also provide the necessary information to detect, model and track objects or people. Furthemore, stereovision is more appropriate for cases where objects or people are very close to the sensor, which is indeed the case for people counting in buses where the integration environment constrainst are very strict. With this technique of stereovision, the objective is to determine the information distance from the sensor to each point of the scene. This distance is inversely proportional to the disparity. We chose dense stereovision instead of sparse one because with this latter the primitives extraction is difficult for the reasons mentioned above. However, even if we chose to exploit dense stereovision, if it is possible we are going to extract specific points like edges. The basic idea of our proposed system is, from disparity maps calculated with a specific stereo-matching, to isolate and separate the passengers' heads in order to count them. To achieve that, the sensor is fixed vertically above the door of the bus. With this configuration, we lower the occlusions problems and whatever the crowd configuration, we suppose that the passengers' heads will rarely touch each other.

3 Processing Chain

The processing chain comprises four blocks: detection with stereo-matching, segmentation, tracking and counting. The detection block calculates, for each pair of stereoscopic images, a dense disparity map which is converted into a height map. On each map are represented distances from the ground of each point of the scene. The height maps are segmented in order to highlight the passengers' heads at different levels (adults, teenagers, children...). The results of this step are binary images containing information related to the heads; we call them "kernels". The extraction block assigns a number of parameters to the kernel: size of the kernel, shape, average greylevel, average height level... Then, with the previous information on the kernels, a tracking procedure is applied to analyze their trajectories.

4 Disparity Maps Calculation

In the literature, for the stereo-matching procedure, there are three classes of techniques of mapping: 1) The global methods: dynamic programming [3], graph theory [4], nonlinear diffusion [5] and spread belief [6]. 2) The local methods: correlation and differential approaches [7]. 3) The cooperative methods [8].

However, the counting application requires a swift technique matching. The SAD algorithm (sum of absolute differences) [9], [10], which is a technique based on the dissimilarity function between the neighborhoods of homologous pixels is an approach that allows a compromise between robustness and processing time. Let us recall that in the context of passengers counting in bus, the sensor is rather close to the observed objects. That is why, we speak about "close stereovision". The main difficulty with this geometrical configuration is the high number of occlusions in stereoscopic images which correspond to regions present in one image and not in the other. To solve this problem, we propose a technique based on a dissimilarity measure called SAD, in which we have added additional constraints.

4.1 Similarity Constraint and Dissimilarity Measurement Weighting

We call similarity constraint any discriminating similarity criterion which may exist between the pixel to match and its homologous or between their respective neighborhoods. The objectives of the use of this kind of constraints are: 1) Reduction of the computation time by rejecting candidates pixels that do not verify the constraints. 2) Increase the accuracy of the matching by choosing pixels strongly correlated.

In our case, we want to improve the stereo-matching quality. Therefore, we propose to refine the selection of homologous pixels by weighing the SAD dissimilarity measure. This is done by introducing a weighting factor whose value depends on the verification of the similarity criterion. When the similarity criterion is verified, the weighting procedures consists in reducing the dissimilarity measure in order to match only the pixels verifying the similarity criterion. The value that the weight can take when the similarity criterion is not verified, does not affect in any way the dissimilarity measure value. For modeling the influence of a constraint, we chose to multiply this similarity criterion by a weighting factor. The dissimilarity measurement is written:

$$C_{SAD}(x, y, s) = \sum |G(x + i + s, y + j) - D(x + i, y + j)| \tag{1}$$

G(x,y): The greylevel of the pixel (x,y) to match belonging to the left image. D(x,y): The greylevel of the pixel (x,y) in the right image. S: The gap between the two pixels (left and right).

The dissimilarity measure after the introduction of the similarity criterion is written as follows:

$$C_{sim}(x, y, s) = coef \sum |G(x + i + s, y + j) - D(x + i, y + j)| \tag{2}$$

with coef, the weighting factor. $Coef = 1$ if similarity criterion is not verified. $Coef = coef_0$ and $0 < coef_0 < 1$ otherwise.

4.2 Proposed Constraints

We have identified four similarity constraints detailed below. Thus, improvement brought by the introduction of the constraints, in terms of accurate matching, does not require a huge increase in the processing time. 1) Similarity of the neighborhoods' centers greylevels. 2) Similarity of the type of pixels (belonging to edges or not). 3) Similarity of greylevels profiles corresponding to the Centerlines of Calculation Neighborhoods (called CCN in the next equation). 4) Similarity of the type of regions including the pixels to match (region with motion or static one). We have respectively associated the coefficients α, β, γ and μ with these constraints. The values of these coefficients will vary depending on whether the constraints are verified or not.

$$\alpha = \begin{cases} 1 \quad \text{if} \quad G \quad \text{and} \quad D \quad \text{do not have the same greylevels} \\ \alpha_0 \quad \text{with} \quad 0 < \alpha_0 < 1 \quad \text{otherwise} \end{cases} \tag{3}$$

$$\beta = \begin{cases} 1 \quad \text{if} \quad G \quad \text{and} \quad D \quad \text{do not represent edges} \\ \beta_0 \quad \text{with} \quad 0 < \beta_0 < 1 \quad \text{otherwise} \end{cases} \tag{4}$$

$$\gamma = \begin{cases} 1 \quad \text{if} \quad \text{the CCN do not have the same greylevels profile} \\ \gamma_0 \quad \text{with} \quad 0 < \gamma_0 < 1 \quad \text{otherwise} \end{cases} \tag{5}$$

$$\mu = \begin{cases} 1 \quad \text{if} \quad G \quad \text{and} \quad D \quad \text{do not correspond to a mobile region} \\ \mu_0 \quad \text{with} \quad 0 < \mu_0 < 1 \quad \text{otherwise} \end{cases} \tag{6}$$

With G corresponding to G(x+s,y), and D to D(x,y). We have determined the optimal values of α_0, β_0, γ_0 and μ_0 experimentally by choosing the values that minimize the stereo-matching error rate according to a given ground truth proposed by Scharstein and Szeliski.

4.3 Constraints Association

So far, we have submitted four similarity constraints that we use to improve the accuracy of matching. Knowing that each of these constraints is of a different nature, it becomes interesting to combine them and analyze their complementarities. The main idea is to gather as much information as possible on the pixel to match and its homologous and also on their neighborhoods. We chose to use an additive model for the calculation of the dissimilarity , which is to add the dissimilarity of the four criteria. The global formulation becomes:

$$C(x, y, s) = (\alpha + \beta + \gamma + \mu) \sum |G(x + i + s, y + j) - D(x + i, y + j)| \tag{7}$$

To evaluate our approach we compared it to SAD (Some of absolute differences), SSD (Some of squared differences), ZSAD (Zero mean some of absolute differences) and ZSSD (Zero mean some of squared differences) on static data with

ground truths proposed by Scharstein and Szeliski and dynamic data with ground truth by using the Vandermark sequence [11]. By comparing the stereo-matching error rate curves of our approach and those of the others we find that our approach has an error rate which is 3% lower than those of the others [12]. This result is mainly obtained in occluded areas of the images. Figure 1 provides two disparity maps calculated on a pair of stereoscopic images. We can notice that for SAD algorithm, some matching errors appear (marked with circles). This shows visually the improvement brought by the introduction of constraints in SAD computation.

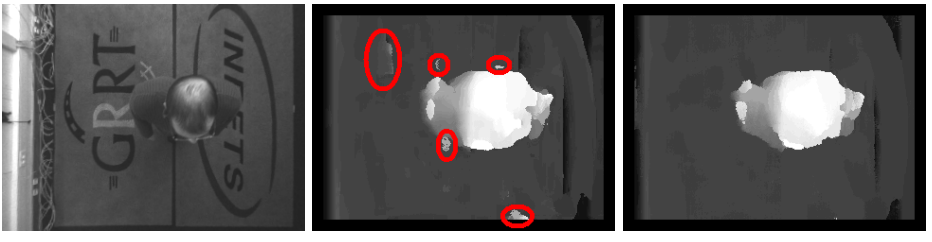

Fig. 1. Disparity maps corresponding to laboratory stereoscopic images (from left to right: initial left image, SAD disparity map with errors outlined with circles, our approach of disparity map)

5 Exploiting Disparity Maps for Tracking and Counting People

The disparity maps calculated by our stereo-matching method are transformed into height maps by simple triangulation. To highlight the passengers' heads, we propose a segmentation technique based on a given number of height intervals. For each interval, we have a binary image on which appear only regions with a height belonging to the interval. In a given interval, the application of morphological operations as openings with circular structuring elements allow us to locate in space and time, the heads of the passengers. At the end of the processing within all the intervals, we obtain a binary image containing kernels supposed to represent the heads. Figure 2 shows an example of segmentation of a height map corresponding to two persons passing under the sensor.

Then, in tracking application, we associate to each kernel a vector of attributes. This vector is a set of properties that distinguish each object from others in the same scene. In our case, the objects we need to track are the passengers' heads and attributes are actually properties that vary from one head to another. The considered attributes are: kernel size, kernel width, kernel length, average height in centimeters, average greylevel (from initial image), kernel coordinates x,y (in the binary image). We use a technique based on Kalman filtering

Fig. 2. Example of a height map segmentation (from left to right: initial left image, the height map, the kernels map)

for the kernels tracking. It yields a prediction of the kernels positions and compares them to the current attributes vectors. When implementing this technique, we have decided to simplify the calculation by limiting the prediction on the two components corresponding to the x and y coordinates of the kernel barycenter in the image. The counting technique we have developed is based on the analysis of the notion of valid trajectories [12].

6 Evaluation and Results

First, we note that the counting system has been fully assessed on real dataset. Data on which the system has been tested come from two different bases: laboratory data according to 30 specific scenario scripts provided by the Parisian Transport Operator (RATP) and 3 real video sequences acquired in an operating bus. Most of the scenarios represent exiting persons. The counting results presented in Figure 3 indicates the number of persons entering or exiting for each sequence of the laboratory. In this figure, we can see the ground truth counting results versus the counting results computed by our algorithm. One can notice that whatever the difficulty of the scenario, the difference between real counting and calculated one with our approach is very low. Indeed these differences are included in the interval $[-1; +1]$. It is also the case for the counting results of the three bus scenarios presented in Figure 4. There are no counting error in the two first scenarios and a slight under-estimation for the third scenario. This is a very encouraging result demonstrating the robustness of our algorithm which is able to cope with various situations (high density groups of persons moving in opposite directions, persons of different sizes, carrying bags...). In order to determine globally the accuracy of our counting system, that is to say considering all the scenarios and mixing entries and exits, we have defined an error rate which is calculated as follows in the formula 8. In this formula, we consider the real counting (ground truth) as basis of comparision and we determine the difference between the counting with our algorithm. Thus, the error rate is around 1%. Among the laboratory scenarios, the same error rate is obtained whatever the

illumination type considered (scenarios 1-15 with daylight and scenarios 16-30 with artificial light). When analyzing more finely the counting results, we see that our system under-estimate systematically the number of persons. Several reasons could explain this fact: 1) the difficulty to detect persons whose size is small and who can be confused with objects. 2) The size of the structuring element in the segmentation step of the disparity map. 3) the merging of two trajectories, corresponding to two different persons.

$$Error_{counting} = 100\frac{(Real_{counting} - Automatic_{counting})}{Real_{counting}} \tag{8}$$

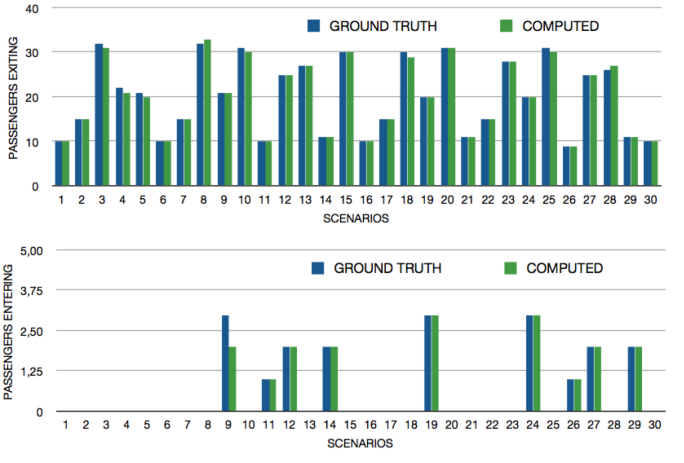

Fig. 3. Counting results for 30 scenarios in laboratory (from top to bottom: exiting from and entering in the bus by the same door)

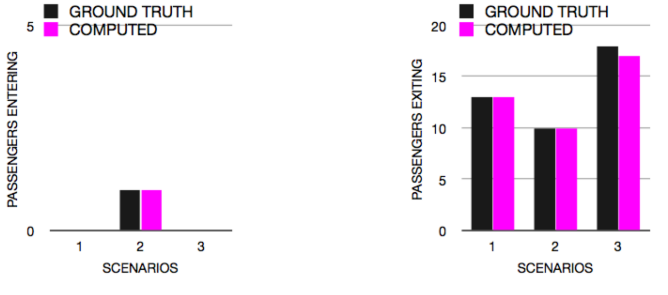

Fig. 4. Counting results for 3 scenarios in bus (from left to right: entering in and exiting from the bus by the same door)

7 Conclusion

In this work, we presented a problem that requires combination of tools emerging from different image processing disciplines. Stereovision was the most important tool we used and adapted to our application. We proposed a stereo-matching approach that integrates similarity criteria to improve the calculation of the disparity maps. It is an issue where the fundamental aspect is attached to the applicative aspect, which allows us to accomplish two objectives, namely, to produce a precise counting system with 99% of good counting and to contribute to the development of new basic tools in the stereoscopic images processing by proposing a matching approach exploiting similarity constraints. The counting accuracy must be refined in a more intensive evaluation. This is currently carried out with a new data set comprising 150 stereo sequences coming from a bus in exploitation. We have good hope to stay under 2% of error thanks to our complete processing chain.

References

1. Beyme, D.: Person counting using stereo. In: Workshop on Human Motion, pp. 127–133 (2000)
2. Terada, K., et al.: A counting method of the number of passing people using a stereo camera. In: 25th IEEE Conference of Industrial Electronics Society, pp. 338–342 (1999)
3. Bobick, A.F., Intillle, H.S.: Large occlusion stereo. International Journal On Computer Vision 33, 181–200 (1999)
4. Paris, S., Sillon, F.: Optimisation a base de flot de graphe pour l'acquisition d'informations 3d à partir de séquences d'image. In: 15 èmes Journées de l'Association Française d'Informatique Graphique, pp. 165–182 (2002)
5. Mansouri, A.R., Mitiche, A.: Selective image diffusion: Application to disparity estimation. In: IEEE International Conference on Image Processing, vol. 3, pp. 284–288 (1998)
6. Sun, J., Zheng, N.N.: Stereo matching using belief propagation. IEEE Transactions on Pattern Analysis and Machine Intelegence 25(7), 787–800 (2003)
7. Wei, G.Q., Brauer, W., Herzinger, G.: Intensity and gradient based stereo matching using hierarchical gaussian basis functions. IEEE Transactions on Pattern Analysis and Machine Intelegence 20(11), 1143–1160 (1998)
8. Zhang, Y., Kambhamatteu, C.: Stereo matching with segmentation-based cooperation. In: 7th European Conference on Computer Vision, vol. 2, pp. 556–571 (2002)
9. Martin, J., Krowley, J.L.: Experimental comparison of correlation techniques. In: International Conference on Intelligent Autonomous Systems (2002)
10. Haris, S., Vandermark, W., Cavrila, D.M.: A comparative study of fast dense stereo vision algorithms. In: IEEE Intelligent Vehicles Symposium, pp. 319–324 (2004)
11. Vandermark, W., Gavrila, D.M.: Real-time dense stereo for intelligent vehicles. IEEE Transactions on Intelligent Transportation Systems 7(1), 38–50 (2006)
12. Yahiaoui, T.: Une approche de stéréovision dense intégrant des contraintes de similarité. Application au comptage de passagers entrant et sortant d'un autobus. PhD thesis, University of Lille (2007)