Fast head turns detection in low quality videos using optical flow

José Mennesson, Benjamin Allaert, Ioan Marius Bilasco

Fast head turns detection in low quality videos using optical flow

José Mennesson[1] Benjamin Allaert[1] Ioan Marius Bilasco[1]

[1] Univ. Lille, CNRS, Centrale Lille, UMR 9189 - CRIStAL
Centre de Recherche en Informatique Signal et Automatique de Lille, F-59000 Lille, France
{jose.mennesson,marius.bilasco}@univ-lille1.fr, benjamin.allaert@ed.univ-lille1.fr

Résumé

Détecter automatiquement une personne qui a besoin d'aide dans un magasin représente un défi en vision par ordinateur. Cela réside dans le fait que la qualité des images est mauvaise. Notre méthode se porte sur un calcul de cohérence des flots optiques denses.

Mots Clef

Cohérence du Flot Optique, Détection de l'Hesitation

Abstract

Detecting automatically a person who needs help in a shopping center represents an unique challenge in computer vision. It lies in the poor quality of images we consider. Our method focuses on consistency of dense optical flow to extract head turns.

Keywords

Optical Flow Consistency, Hesitation Detection

1 Introduction

In a shopping center, detecting persons who are lost, searching a product or in general, need help represents important tasks to maximize the efficiency of the sales assistants. A customer is considered as hesitant when he is uncertain where to go. In this case, we assume that a wandering customer has a high head-turning frequency. The field of recognizing human behavior is one on the main topics of video analysis. In this domain, we study the head pose estimation problem which remains a challenge in computer vision. For head pose estimation in near to frontal situation, neural network, manifold learning, geometric, tracking methods and so on can be used [?]. However, these methods make some assumptions to perform well such as continuous video, frontal view, familiar-identity assumptions and so on. In our context, estimating the head pose is not required since only the head movements (turn right/left) are necessary. Among the approaches cited above, methods based on optical flow can be used to obtain accurately and in real-time the movements of face even when the head image is noisy and degraded. Optical flow represents the pattern of apparent motion of objects in a video caused by the relative motion between the camera and the scene. When good quality videos are considered, head movements can be estimated accurately as in [?]. The most common method to estimate the optical flow is the Lucas and Kanade one [?]. When the extraction of interest points is not feasible (for example, low-quality video), then a dense optical flow can be estimated as proposed by Farnebäck in [?]. In this paper, we propose to use the optical flow inconsistency induced by head turns to detect it.

2 Context

In our context, a network of several cameras can be used to detect wandering customer. In this paper, it is assumed that an hesitant person yield high head-turn frequency. To detect head-turn within a camera network, only left and right head turns of customers are considered. Actually, four cameras are installed horizontally with respect to the ground at the four corners of the room, so horizontal movements of the head will be captured by the cameras as horizontal movements. Hesitation detection can be performed on each camera and each decision (turn/not turn) can be aggregated (e.g. by using a majority vote) to improve the results. In this paper, we assume that the customer is detected and re-identified through the camera network as in [?].

3 Overview of our method

First, persons are detected, re-identified and heads are extracted from the key frames of the four videos. Then, dense optical flow is computed between consecutive heads for each camera and the percentage of consistent vectors is used as feature to detect head turns. Finally, to merge the results provided by cameras, each decision is re-synchronized by propagating results from key frames to adjacent duplicated frames (see Figure 2). The final decision can be obtained by using a majority vote.

- Dense optical flow corresponds to an optical flow which has been computed on a dense grid of points. This method is convenient when it is difficult to extract keypoints (e.g. a picture of a face). The most popular algorithm which was developed to produce dense optical flow is the Gunnar-Farnebäck algorithm [?]. In our context, the challenge lies in the poor quality of head pictures (see Figure 1) which implies that the optical flow is not necessarily accurate.

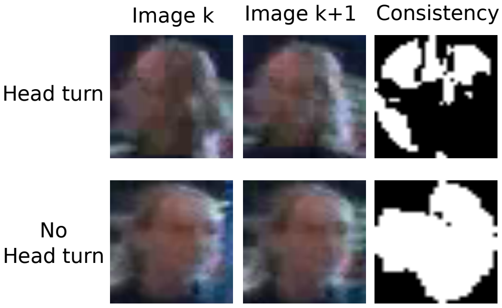

FIGURE 1 – Results of consistency computation.

- Optical flow consistency : most of optical flows methods are not robust to many factors as occlusions, noise, huge movements, etc. One can try to correct false flows (increase computing time) or one can consider only consistent flows. As head turns induce occlusions, which produce inconsistent flows, we consider the percentage of consistent flows to detect head turns.
To test the consistency of an optical flow, it is computed two times : between the current and the following frame (forward flow) and between the following and the current frame (backward flow). If a vector of the forward flow is the exact opposite of the backward one, then it is considered as consistent. This is computed as :

$$consistency(\mathbf{v_f}, \mathbf{v_b}) = \begin{cases} 1 & \text{if} \quad \measuredangle(\mathbf{v_f}, \mathbf{v_b}) < \theta_{eps} \\ 0 & \text{else} \end{cases}$$

where $\mathbf{v_f}$ (resp. $\mathbf{v_b}$) correspond to optical flow vectors computed forward (resp. backward). Example of consistency results on head turn is shown in Figure 1. The considered feature is the percentage of consistent vectors in optical flow computed on segmented heads pictures. If this feature is lower than a threshold (c_{thres}) fixed empirically by the user, then we consider that there's a head turn.
- Merge results : in order to synchronize videos, some duplicate frames are added by the system between the key frames as it is shown in Figure 2. In this case, optical flow can't be computed directly between two consecutive frames because the optical flow would be null. So, optical flow consistency must be computed between two consecutive key frames and the results can be propagated to the following duplicate frames for each camera. Then, results can be aggregated by using a majority vote of the cameras.

4 Experiments

In this paper, we have used a non-public video surveillance dataset collected by ©NOLDUS company during ITEA2 Empathic project. Three scenarii taken by 4 video cameras are considered in this paper. Each scenario corresponds to one person who walks and turns its head quickly (FHT) or slowly (SHT) several time. Each segments lasts between 3 and 7 seconds. Heads are detected manually and segmented using Grabcut algorithm [?]. Head turns detection is done frame by frame. In this paper, $\theta_{eps} = 10^{\circ}$and $c_{thres} = 0.6$. Results in terms of precision and recall are shown in Table 1. Our method works well for Scene2 where the person turns his head quickly. Underachievement of Scene1 and Scene2 are due to sudden changes in the background (which induces inconsistent optical flow) and slow head turns which are hard to detect by optical flow.

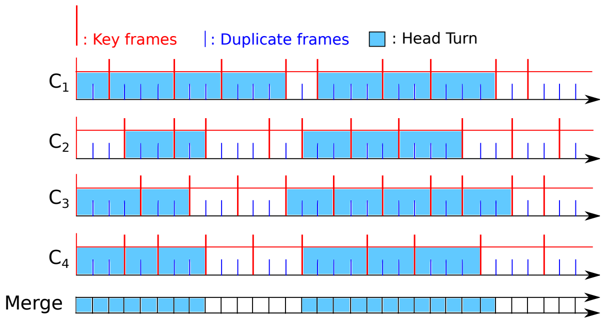

FIGURE 2 – Decision merging from 4 cameras decision

Results in %	Precision	Recall
Scene1 (2 FHT & 1 SHT)	0.57	0.73
Scene2 (2 FHT)	0.95	0.81
Scene3 (2 FHT & 2 SHT)	0.37	0.67
Mean	0.58	0.74

TABLE 1 – Results for 3 scenarii.

5 Conclusion

This preliminary work shows that dealing with the consistency of optical flows can provide information when accurate flow can't be obtained. This method is faster than other methods which propose to overcome optical flow limitations.

Acknowledgements

This work was supported by "Empathic Products", ITEA2 1105.

Références

[1] G. Farnebäck. Two-frame motion estimation based on polynomial expansion. In J. Bigun and T. Gustavsson, editors, *Image Analysis*, LNCS, pages 363–370. Springer, 2003.

[2] B. D. Lucas and T. Kanade. An iterative image registration technique with an application to stereo vision. IJCAI'81, pages 674–679, CA, USA, 1981.

[3] S. Martin, C. Tran, A. Tawari, J. Kwan, and M. Trivedi. Optical flow based head movement and gesture analysis in automotive environment. In *ITSC*, pages 882–887, 2012.

[4] E. Murphy-Chutorian and M.M. Trivedi. Head pose estimation in computer vision : A survey. *PAMI*, 31(4) :607–626, 2009.

[5] C. Rother, V. Kolmogorov, and A. Blake. Grabcut : Interactive foreground extraction using iterated graph cuts. In *ACM transactions on graphics*, volume 23, pages 309–314, 2004.

[6] R. Zhao, W. Ouyang, and X. Wang. Learning mid-level filters for person re-identification. In *CVPR*, pages 144–151, 2014.