MULTI VIEW PERSON LOCALIZATION

Ing. Stanislav SUMEC, Doctoral Degreee Programme (3)
Dept. of Computer Graphics, FIT, BUT
E-mail: sumec@fit.vutbr.cz

Supervised by: Dr. Pavel Zemčík

ABSTRACT

This paper describes a technique how persons can be localized in 3D space using several cameras. Various methods are used to person detection in pictures obtained from cameras. Position of person is then identified from positions of objects detected on different cameras with overlapping visual angles and mutual positions of these cameras. Proposed method is used as auxiliary technique for an identification of meeting participants.

1 INTRODUCION

Person localization can be useful in many cases. One of its possible applications is a securing of protected areas such as buildings. It is common that such areas are monitored by several cameras. Modern technology can help with processing of obtained data and can also save human work. For example motion detection can be used to operator warning that something is moving in watched areas. It is possible to use various known image processing algorithms for detection of moving objects and humans on cameras pictures [5][6]. These algorithms make possible to obtain position of detected objects in 2D space of recorded pictures. But knowledge about position in 3D space of given area can be sometimes useful. This paper describes one possible method for an evaluation of person location using pictures from several cameras. The main idea of the algorithm is that the desired object is detected on more than one camera with overlapping views and its localization is deduced from mutual positions of used cameras and detected positions. At the first the methods for object detection in pictures are described. Further the main idea of localization algorithm is sketched. Current application of the proposed algorithm in labeling of meeting participants is mentioned is at the end of the paper.

2 DETECTION

The first necessarily thing for person localization is detection of person's 2D positions in pictures obtained from all cameras. It is possible to use various methods for this purpose. For example skin color detection can be used for detection of human's parts with visible skin such as faces or hands. The specific color and probability models are used in this technique [2]. A little disadvantage of this method with regard to human detection for localization is that the human is detected only if his face or hands are directed to the camera.

Other possible method uses a reference picture obtained from the camera when nobody occurs at its view. Humans can be detected so that current picture and the reference picture are subtracted. Following equation shows how can be computed value expressive difference of pixel in source image *S* and reference image R. Constants *r*, *g*, *b* are weights of particular color component.

$$D(x,y,t) = r\left|S_r(x,y,t) - R_r(x,y)\right| + g\left|S_g(x,y,t) - R_g(x,y)\right| + b\left|S_b(x,y,t) - R_b(x,y)\right| \quad (1)$$

The threshold is applied to obtained values to distinction if some new object occurs in camera pictures. An advantage of this method is that whole human is detected not only parts with visible skin. So it is possible to detect person from an arbitrary angles. But obtaining of references pictures is limiting for this method so it can be use mainly for cameras with fixed view.

Next method is based on differences computation between two or more following frames. Number of processed frames is given by constant *N*.

$$D(x,y,t) = \frac{1}{N}\begin{pmatrix} r\sum_{i=0}^{N-1}\left|S_r(x,y,t-i) - S_r(x,y,t-i-1)\right| + \\ g\sum_{i=0}^{N-1}\left|S_g(x,y,t-i) - S_g(x,y,t-i-1)\right| + \\ b\sum_{i=0}^{N-1}\left|S_b(x,y,t-i) - S_b(x,y,t-i-1)\right| \end{pmatrix} \quad (2)$$

Source picture | Skin color detection

Reference picture detection | Motion detection

Fig. 1: *Person detection*

Threshold is also applied to resulting values. This technique enables to detect moving objects on camera pictures. But it can be sometimes insufficient because immovable humans can not be detected. Each of mentioned methods has some advantages and disadvantages so the best results can be achieved using combination of several possible methods. Experimental results obtained with described methods can be seen in the figure 1. Segmentation is applied to obtained pictures to separating persons from the rest of the visual image.

3 LOCALIZATION

If some objects are detected in camera picture their position in 3D space can not be still evaluated. But the set of possible localization can be eliminated to the positions at a ray outgoing from camera. Direction and begin of the ray is given by 2D position of detected object in camera picture. Used geometry is the same as a visualization algorithm Ray casting uses [3]. It is assumed in this algorithm that color of a pixel at given position in the picture is computed from color of the nearest object intersecting the ray emitted for this pixel. Properties of the rays can be computed according to preset parameters of camera as focal length, its position and direction. It is also possible to use calibration according to positions of known objects, which occurs in camera view. For example a screen can be seen in figure 1. Its position in camera picture can be detected and together with its known position in 3D space parameters for ray computation can be obtained.

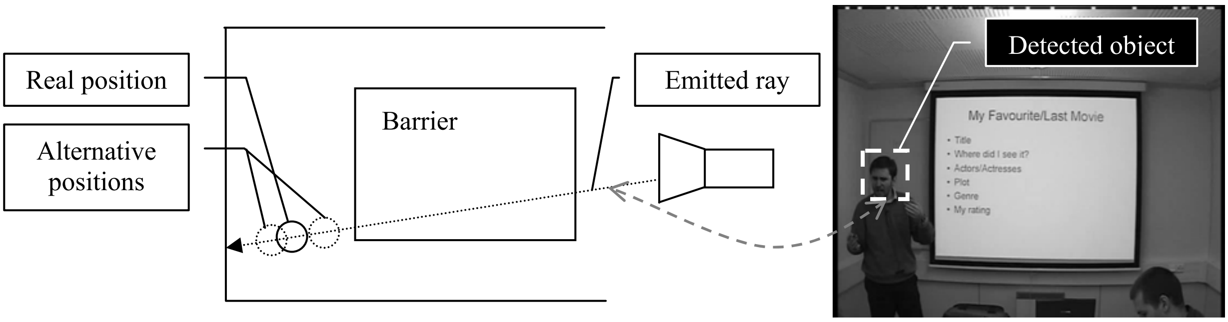

Fig. 2: *Localization with one camera*

Figure 2 shows a real situation in the meeting room where the head of one person was detected. Corresponding ray was emitted according to position of detected object to determination of person's position. As was mentioned all positions on the ray are possible. Fortunately knowledge about meeting room setup can help with person localization, because possible positions can be eliminated only to areas, where the person can walk. But one camera is still not enough for more precise localization. The main idea of proposed solution is to use at least two cameras watching the destination area from different overlapping visual angles. If desired object is detected on several camera pictures corresponding rays can be computed. The resulting position in 3D space is theoretically given by intersection of the rays emitted from every camera. But minimal distances among the rays have to be found in fact. The minimal distance of two rays given by equations ${}^{1}p = A + \vec{q}r$ and ${}^{2}p = B + \vec{s}t$ is

$$d = \frac{\left\|\left[(B-A)\vec{q}\vec{s}\right]\right\|}{\left|\vec{q}\times\vec{s}\right|}. \tag{3}$$

The resulting position is then given by abscissa, which are joining the rays with minimal distances. Figure 3 shows principle of localization with two cameras.

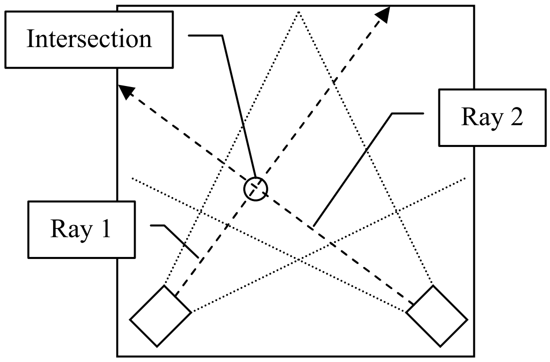

Fig. 3: *Localization with two cameras*

Limitation of this technique is that if watched person is not located in the area where the camera visual angles overlap, localization can not be evaluated. The solution of this problem is felicitous placement of cameras such way that whole monitored area is covered at least two cameras. Usage of knowledge about the watched spaced is also advantageous, because impossible positions can be directly excluded.

4 APLICATION

The proposed technique for person localization in 3D space together with other techniques is currently used for a labeling of meeting participants. Source meetings are recorder by three cameras with partly overlapping visual angles. The problem is to assign identification to every detected participant. The identification is not only assigned to whole person but heads and hands of meeting participants are detected and identified separately. This identification has to hold during whole meeting even if the participant walks through the meeting room. Preservation of person identification, when the person is walking from one camera to other camera, is especially problematic. Identification has to work also when more than one participant is walking through the room. The overlapping visual angles of the

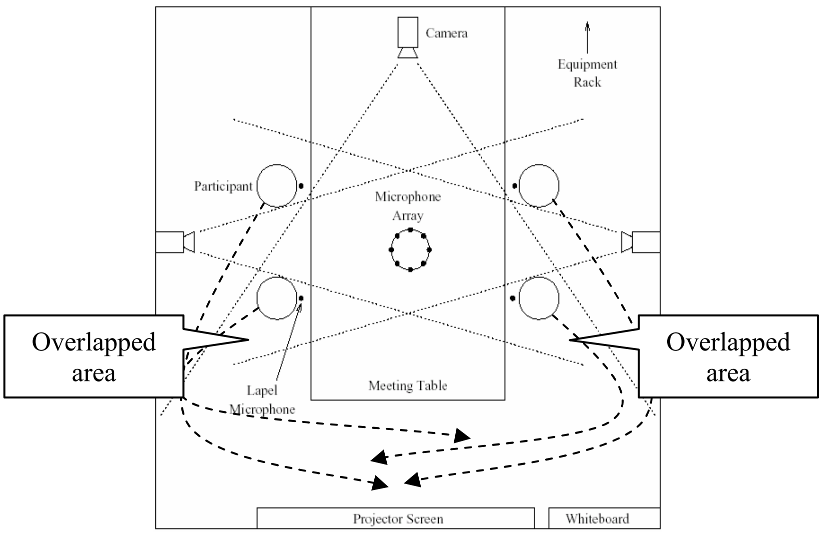

Fig. 4: *Meeting room setup*

cameras can help with this problem. Used meeting room setup can be seen in the figure 4. Areas with overlapping camera view angles and possible traces of participants are highlighted. As can be seen if somebody is walking from one camera view to other camera view always passes through the overlapped area. The principle of used technique is following. It is supposed that given participant is detected and identified on certain camera. The ray outgoing from this camera can be computed for detected position and so possible positions in 3D space can be obtained. When the person is walking through the meeting room and new person is detected on other camera, new ray for this person is computed. This new ray defines possible positions of new detected person. If the minimal distance between the ray obtained from original camera and the ray representing new position on other camera is less than given threshold identification of this person can be assigned to the new detected person. But other techniques are also used for better results achievement [4]. Proposed methodology was experimentally verified and obtained results shown that it is usable.

5 CONLUSION

The method for person localization using several cameras was presented. As was shown described principles can be used besides localization also for preserving of meeting participant's identification when the participant is walking from one camera view to other camera view. Big advantage of using described method for this purpose is its low computation requirements, because only simple geometry model has to be evaluated. It is also possible to use other methods for example face recognition, but their computation consume much more time.

ACKNOWLEDGEMENTS

The paper has been founded by EU IST Program project IST-2001-34485. It is part of CPA-2: the Cross Programme Action on Multimodal and Multisensorial Dialogue Modes. For more information refer to [1].

REFERENCES

[1] The MultiModal Meeting Manager (M4) Project (available at http://www.dcs.shef.ac.uk/spandh/projects/m4)

[2] Potúček, I.: Person Tracking Using Omnidirectional View, In proceedings of EEICT 2003, Brno

[3] Watt, A.: 3D Computer Graphics, Addison-Wesley, Wokingham, UK, 1993

[4] Potúček, I., Sumec, S.: Participant Activity Detection by Hands and Face Movement Tracking in The Meeting Room, To appear in CIG 2004 Conference Proceedings, Greece

[5] Aggarwal, J. K., Cai, Q.: Human Motion Analysis: A Review, IEEE Nonrigid and Articulated Motion Workshop, June 1997

[6] Kruppa, H., Bauer, M., Schiele, B.: Skin Patch Detection in Real-World Images, Perceptual Computing and Computer Vision Group, ETH Zurich, Schwitzerland, 2002