Marker-less Vision Based Tracking for Mobile Augmented Reality

D. Beier, R. Billert, B. Brüderlin
Technische Universität Ilmenau
Computer Graphics Program
Ilmenau, Germany
david.beier@tu-ilmenau.de

D. Stichling, B. Kleinjohann
Universität Paderborn
Intelligent Mobile Systems
Paderborn, Germany
tichel@c-lab.de

Abstract

In this paper an object recognition and tracking approach for the mobile, marker-less and PDA-based augmented reality system AR-PDA is described. For object recognition and localization 2D features are extracted from images and compared with a priori known 3D models. The system consists of a 2D graph matching, 3D hypothesis generation and validation and an additional texture based validation step.

1. Introduction

Augmented Reality (AR) is gaining importance in industrial applications, for developing, production and servicing but also for mobile applications [2]. In the AR-PDA application no external tracking devices are provided, therefore 2D features extracted from the camera images are used. Because the objects in the scene and their shapes are known a model-based approach is used.

The object recognition process consists of two parts. During the first part, *2D Vision*, the images are analyzed and straight edges are extracted, vectorized, and matched with 2D views of known 3D models. In the *3D Vision* part hypotheses about possible combinations of features are generated, verified, and the best matching hypothesis is chosen. For final validation of the results image-based rendering techniques are used.

2. 2D Vision

The goal of the 2D vision is to extract 2D features of the objects to be searched. Due to the fact that AR-PDA deals with nearly rectangular scenes straight edges are used as characterizing features. These extracted and vectorized edges are matched with *2D views* of the 3D object models.

The pixel images are preprocessed using a *Sobel Filter* and a *Non-Maxima Elimination* and finally vectorized using an algorithm proposed in [3]. The vectorized edges are virtually elongated and the intersections of the elongated edges are used as vertices of a 2D graph describing the edge structure of the input image called the *2D image graph*.

2D views of the 3D model are used to identify potential locations of the objects to be searched. These 2D views consist of two types of edges: *essential* and *optional* edges. Using graph-matching algorithms potential locations of the 2D views inside the 2D image graph are evaluated. The constraints for the matching algorithm are the aspect ratios between the length of the edges and the angles between the edges.

Figure 1 shows the four steps of the 2D vision: a) the vectorized edges, b) the virtually elongated edges, c) one match of the essential edges, d) one match of the 2D view including neighboring edges.

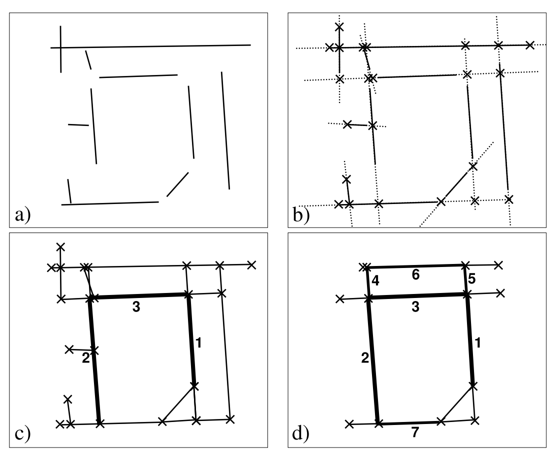

Figure 1. Generation of the 2D graph

3. 3D Vision

In the 3D vision part the 2D features are compared with the database of 3D-models. The models are based on sim-

Proceedings of the Second IEEE and ACM International Symposium on Mixed and Augmented Reality (ISMAR '03)
0-7695-2006-5/03 $17.00 © 2003 **IEEE**

plified CAD data, containing highly visible edges, faces and texture information.

With correspondences of image features and 3D-model features hypotheses for the orientation of the model relative to the camera are generated. Each generated hypothesis will be verified by projecting the model into the image plane. This projection is compared with the extracted edge graph of the input image and the matching of both graphs is evaluated. The best matching hypothesis is taken to determine the recognized object, its location and orientation relative to the camera coordinate system. This information is needed for augmentation.

Because the main objective of the AR-PDA application is the recognition of technical objects with many straight edges, edge correspondences are used to compute the orientation of these objects. With three 3D–2D–edge-pairs it is possible to compute the relative orientation of the camera to the object [1], therefore hypotheses consisting of three edge-pairs are used.

The approach by Dhome et al. [1] has been implemented, which finds the analytical solutions by solving an equation system for inverse perspective projection. It uses a simplification of the pose determination problem by implicitly dividing it into lower dimensional subproblems.

Testing of all possible hypotheses has polynomial-time complexity. Therefore the number of hypotheses has to be decreased as much as possible. To do this the following strategies are used: *Reducing the number of edges*, *Assuming limited camera orientation*, *Classification of edges*, and *Using invariant relations*.

The goal is to find the correct hypothesis as fast as possible. Therefore the system tries to find more probable hypotheses first. This way, the number of hypotheses is reduced because a suitable solution is found more early.

4. Recognition of Variations of Objects

For different objects with identical or similar geometry, it is difficult to recognize these objects only by using the geometry features such as edges.

For recognition of variations of objects (e.g. models of the same brand), texture matching of regions of interest is used. These regions of interest have to be separated and normalized first. Because the base model and its orientation in the image is known from the previous processing step of hypotheses generation, the location of the region of interest can be determined. Usually, these regions are distorted by perspective projection in the image. So with the known extrinsic camera parameters, the regions are deskewed and normalized.

On the normalized regions filter operations for comparing these regions with known models are applied. Especially a fast histogram matching approach to compare the deskewed regions with the models has been realized.

For the deskewed image and for every model histograms of the vertical gradients along the x axis are computed. The histograms of all models are saved beforehand in a grayscale image (see Figure 2). The extracted histogram is compared by cross correlation with the reference histograms and the best matching correspondence identifies the model.

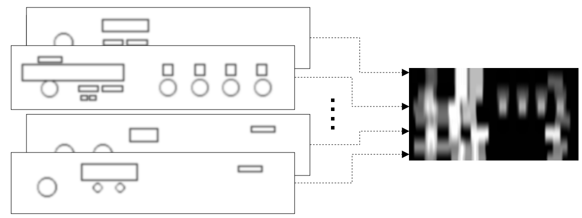

Figure 2. *Grayscale histograms of different models*

5. Results

The algorithms are implemented on a PC-based server communicating with a PDA using Wireless LAN resulting in a framerate of about 6-10 frames per second.

6. Conclusion

The computer vision algorithms of the AR-PDA project have been presented, proving that it is possible to realize a fast marker-less and vision based AR tracking system for mobile systems.

7. Acknowledgment

The AR-PDA project is funded by the German Federal Ministry for Education and Research (BMBF) under grant no. 01 IRA 13F.

References

[1] M. Dhome, M. Richetin, J.-T. Lapreste, and G. Rives. Determination of the attitude of 3-d objects from a single perspective view. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11(12):1265–1278, 1989.

[2] C. Geiger, B. Kleinjohann, C. Reimann, and D. Stichling. Mobile AR4ALL. In *Proceedings of The Second IEEE and ACM International Symposium on Augmented Reality (ISAR 2001)*, 2001.

[3] D. Stichling and B. Kleinjohann. Edge vectorization for embedded real-time systems using the CV-SDF model. In *VI2003: 16th International Conference on Vision Interface*, Halifax, Canada, June 2003.