VIRTUAL ADS INSERTION IN STREET BUILDING VIEWS FOR AUGMENTED REALITY

Yu Huang†, Qiang Hao‡, Heather Yu†

Huawei Technologies USA†, West Virginia University‡

ABSTRACT

This paper is proposing a new framework of virtual ads insertion for Augmented Reality (AR) in a specific real scene, building facades of the street view videos. In this framework, specific region detection, camera orientation estimation, visual tracking, motion filtering and virtual-real blending are discussed. For the building façade, vanishing points are fully identified to detect a rectangular planar structure. Meanwhile, dynamic registration and color harmonization are designed to improve visual acuity of virtual ads insertion. Experimental results are given to show the good performance of the proposed framework.

Index Terms— Augmented reality, virtual ads insertion, dynamic registration, vanishing point, color harmonization.

1. INTRODUCTION

Augmented Reality (AR) is getting closer to real world consumer applications. The user expects the augmented content to better comprehend and enjoy the real scene, like the sightseeing, sports game and work place. One of its applications is ads insertion, also being a category of virtual content insertion (VCI) [6], which can help broadcasters and content distributors to increase the additional revenue. The basic concept consists of identifying specific places in the real scene, tracking them and augmenting the scene with the virtual ads. Specific region detection relies on scene analysis and understanding.

For virtual ads insertion, the challenging issues can be listed as: how to less intrusively insert the contextually relevant ads (what) at the right place (where) and the right time (when) with the attractive representation (how) in videos. Liu et al. proposed a general VCI system [6] which performs attention analysis to detect the higher attentive shot as the insertion time and lower attention region as the insertion place. Unfortunately structure information is not utilized to find the meaningful object for ads augmentation. An approach to insert virtual ads in the video was proposed [8] where they found and segmented planar surfaces in the scene automatically as the insertion places. Obviously the state-of-art segmentation techniques cannot prove good performance in planar surface recognition. Medioni et al. proposed to replace the billboard in the scene [11]. However, it may cause legal issues due to original ads substitution.

Sports video ads insertion has been under research for its large commercial value. Generally advertisers would hire professional editors to manually implement, however it may be very labor-intensive and inefficient for rapid productions on monetizing sports video in this way. Compared to a general system [6, 8], automatic ads insertion in sports video looks easier for the well understanding of domain knowledge. For example, Wan et al. [11] selected the region above the goal mouth bar or the soccer central ellipse for placing virtual ads. It can be seen playfield can be modeled and used for foreground (players) separation. Chang et al. applied tennis court model fitting and tracking to dynamically insert ads [2], in which visual acuity was analyzed and color harmonization was performed to reduce disturbance of virtual ads to the viewers.

For street view videos, building façades are mostly observed. Inserting virtual ads into this kind of videos also shows a big potential market. Especially, building or street recognition [5] for AR provides the clue for relevant ads selection. If we know who is watching (registered users) or capturing (via a mobile device) the video, targeted ads is enabled to be much more appealing. Alvarez et al. proposed an interactive system that allows inserting a virtual picture in the architectural image [1]. This tool needs a few interactions for users to select two groups of parallel lines for both of vanishing point identification and camera calibration and to mark manually a rectangular region for content insertion. It didn't address how to insert new elements in the video either.

We propose a new framework of automatic virtual ads insertion in the street view videos, i.e. inserting ads on the building façades. First, we extract straight lines from the image and then identify the vanishing points. From two sets of parallel lines corresponding to respective vanishing points, we detect a dominant rectangular planar structure that satisfies corner verification and dominant direction verification for ads insertion. To realize seamless effect of ads insertion in the video, we come up with a close loop of detection-and-tracking for dynamic registration. Especially, motion filtering is tailored to reduce jittering in the virtual-real alignment for AR. Finally, to make the viewers of the augmented video much less disturbed, color harmonization is tuned for virtual-real blending. Experiments and results are given to demonstrate the good performance of the proposed framework.

2. ALGORITHMS

Figure 1 shows the diagram of the proposed framework for virtual ads insertion in the street building scenes for AR. This framework is also feasible for other kinds of scenes, like soccer and tennis videos [2, 11]. In this paper, specific region detection includes vanishing point estimation from parallel straight line detection and dominant rectangular planar extraction, camera calibration and model fitting means homography estimation of 3-D planar patch and virtual ads image plane, and visual verification of tracking results runs by corner verification and dominant direction verification in ads insertion.

2.1. Vanishing Point Estimation

To avoid representing edges on the Gaussian sphere, we apply a non-iterative approach for vanishing point estimation [10] with slight modification, based on a recently proposed algorithm for the simultaneous estimation of multiple models called J-linkage.

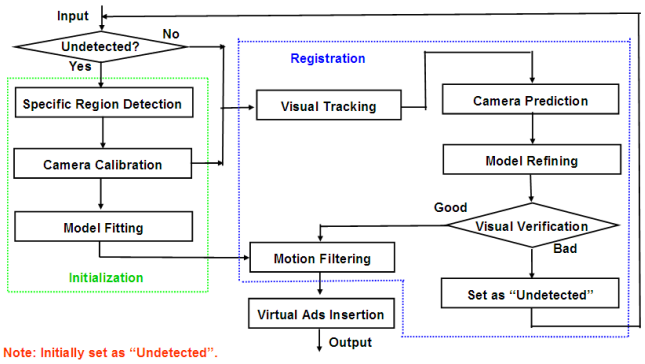

Figure 1. Diagram of virtual ads insertion in real scenes

First, the Canny edge detector followed by non-maximal suppression is used to obtain a map of one pixel thick edges. Then, junctions are eliminated and connected components are linked using flood-fill. Each component is then divided into straight edges by browsing the list of coordinates. It will split when the standard deviation of fitting a line is larger than one pixel. We also merge separate short segments which lie on the same line to reduce error and also to reduce computation complexity in classifying lines. Figure 2 is an example of edge map in red.

Figure 2. Edge map of a building façade

The vanishing point estimation algorithm [10] is based on the J-linkage that requires a measure of consistency between hypothesized vanishing points and edges that is geometrically meaningful, treats all vanishing points equally and does not require distinguishing finite and infinite vanishing points. The objective function is minimized as

Figure 3. Vanishing point estimation

$$\hat{v} = \arg\min_{v} \sum_{\varepsilon_j \in S} w_j^2 dist^2([\bar{e}_j]_\times v, e_j^1) \quad (1)$$

with e_j^i as two end points of edge ε_j, $\bar{e}_j$ their centroid, v the vanishing point and w_j the weight. Eventually EM is performed to refine the solution generated by the J-linkage. Figure 3 is an example of vanishing point estimation, shown with a group of parallel lines in green.

2.2. Rectangular Planar Structure Extraction

Rectangular planar structure is an image of a 3D rectangle. Structures formed by two pairs of parallel lines on a plane corresponding to orthogonal vanishing points might be rectified into rectangular patches. This assumption works for most man-made environments, like the building facades in the street view scenes. However, there are too many hypothesis generated from mutually orthogonal parallel lines. A refinement approach of merging, pruning and verifying the rectangular structure hypothesis [12] is applied to find a biggest rectangular patch for ads insertion.

First, we merge the segments lying on the same line and suppress lines those are either close-by or too short. Moreover, both the line candidates are sorted from left to right or from top to bottom. Figure 4 is the edge map (in green) before and after the line refinement process.

(a) Before line refinement

(b) After line refinement

Figure 4. The edge map of a building facade

Two observation truths are used for rectangular planar structure hypothesis verification [12]: 1) The four intersections must be actual corners in the building facade, and it deletes the case of intersections of lines in the sky; 2) the front-view of this

image patch must only contain the horizontal and vertical directions. To realize that, we need to recall that any planar mapping between the 3-D world plan and the image plane can be characterized by a homography [4], which relates the coordinates of points from two respective planes. Represented in homogeneous coordinates, the homography transformation $p = Hp'$ is rewritten as

$$\begin{pmatrix} x \\ y \\ w \end{pmatrix} = \begin{pmatrix} h_{00} & h_{01} & h_{02} \\ h_{10} & h_{11} & h_{12} \\ h_{20} & h_{21} & h_{22} \end{pmatrix} \begin{pmatrix} x' \\ y' \\ w' \end{pmatrix} \quad (2)$$

Homogeneous coordinates are scaling invariant, reducing the degrees of freedom of H to only eight. Actually four extreme points of a rectangular structure are sufficient to estimate the homography matrix that enables us to warp the hypothesized image patch to a normalized fronto-parallel view. Then the gradient histogram is calculated to find the dominant directions. Figure 5 shows a deleted rectangular patch (in green) and its front-parallel view because there are additional peaks in its gradient histogram. Figure 6 gives a detected dominant rectangular planar region (in blue).

(a) A rectangle hypothesis (b) Frontal-view

Figure 5. A false hypothesis of rectangular structure

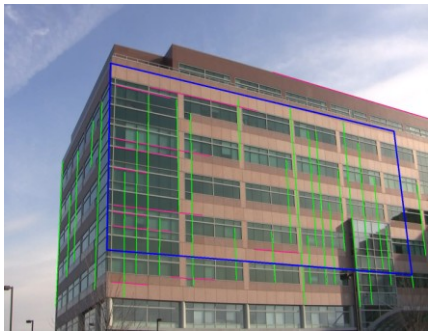

Figure 6. A detected rectangular structure

2.3. Virtual Ads Insertion and Color Harmonization

When we warp the rectangular planar patch into its fronto-parallel view, the virtual ads frame can be scaled to replace this patch and then is warped back to the original plane in the real scene. However, the representation of projected ads plays an important role in attaining realistic object assimilation.

A color harmonization method [2] is applied to ensure the advertising effect being effective and less intrusive, which means the color scheme of the ads frame is adjusted for providing visual aesthetics and the sense of harmony according to empirical harmony theories of colors. In this case, the inserted ads frame is in harmony with the building façade by a modified alpha-blending method as

$$I'(x,y) = (1-\alpha)I(x,y) + \alpha I_{Ad}(x,y), \quad (3)$$

where $I(x,y)$, $I_{Ad}(x,y)$ and $I'(x,y)$ is the original image value, ads value and the actual inserted value at pixel (x,y), and α is the normalized opacity. Details of how to estimate α by a contrast sensibility model is referred to [2]. Figure 7 gives examples of ads insertion with/without color harmonization (the former makes insertion look more natural).

Figure 7. With/Without color harmonization

2.4. Dynamic Registration by Detection-and-Tracking

Dynamic registration is a challenging problem in AR, so how to dynamically register the inserted virtual ads with the replaced or overlapped rectangular planar patch is critical to realize a good performance of visual acuity. As a matter of fact, it is a tracking problem, visually only in this case. Tracking based on sensors is relatively easy, available for mobile devices. Visual tracking is much more accurate than sensor-based; however it may easily fail during rapid camera movement, i.e. the drifting artifact. Here we propose a dynamic registration method with a close loop of detection-and-tracking, where visual detection is used for rebooting the tracking process to avoid drifting error and motion filtering is performed to reduce jittering (Figure 1).

The tracking method is based on the popular keypoint-based KLT method [9]. An example of keypoint tracking is given in Figure 8 where those trajectories of key points (in blue) on the rectangular patch are drawn in yellow.

Figure 8. KLT-based keypoint tracking

Based on the tracked keypoints inside the rectangular planar patch, we will update the homography matrix and the rectangular patch locations. The estimation method should be robust to outliers, like RANSAC. We apply two hypothesis verification rules in session 2.2 to decide if the tracking result is good, i.e. whether the corners of this rectangular structure within the image are still the corners of the building façade and the gradient histogram of its fronto-parallel view does not own additional peaks except the horizontal and vertical directions.

Besides, motion filtering is necessary to smooth the estimated camera motion parameters because tiny errors in KLT

tracking may cause frequent switching between detection and tracking and jittering in ads insertion. We apply a Wiener filtering method for smoothing the tracking result with buffers [3] once the detection process triggers the tracking process. We assume the inserted patch's corner locations p_i^j (j=1~4) in the current ith frame are linear combinations of those in previous N and following N frames:

$$p_i^j = \sum_{k=-N}^{N} \beta_{i+k} \cdot p_{i+k}^j \quad (4)$$

Then the $2N$+1 optimal coefficients are estimated with training samples by a LS formulation [3]: assume the number of buffer is M, then there are M-$2N$ training samples; A data matrix C with size (M-$2N$)x($2N$+1) and a sample vector $\bar{p}$ with size (M-$2N$) are built to estimate the optimal coefficient vector $\bar{\beta}$ with size ($2N$+1) by minimizing $\| \bar{p} - C\bar{\beta} \|^2$.

3. EXPERIMENTAL RESULTS

Figure 9. Ads insertion for building A

We have captured some HD (1440x1080) street videos with a hand held camcorder for ads insertion experiments. To be simple, the logo of Huawei Company is selected for insertion, seen in Figure 7(b). Some ads insertion results are shown in Figure 9 (frame # 99, 199, 299, 399, 499 and 569 in video A) and Figure 10 (frame # 149, 299, 599, 749, 899 and 1129 in video B) from two videos of respective building scenes. It is seen the virtual-real alignment and visual acuity looks satisfying even with slight hand shaking. In Figure 10, occlusion by trees is not handled, however color harmonization reduces the annoying effect. The process is not real-time, since detection of rectangular structure takes about one second per frame.

Figure 10. Ads insertion for building B

It is also possible to insert a video on the real scene, for example given in Figure 11(a). Besides, our method can be tailored to apply for tennis court videos too, except the model fitting is constrained to a tennis court scenario, illustrated in Figure 11(b). It can be seen the occlusion handling is easy by playfield modeling in advance.

4. CONCLUSIONS

In this paper, we have implemented a novel framework of ads insertion for AR. To realize a good performance of seamless virtual-real blending, we applied a close loop of detection-and-tracking plus motion filtering for dynamic registration and color harmonization for visual acuity.

In future, we intend to improve the virtual-real alignment by salient feature matching (SIFT in MSER) and also add modules to handle some critical issues in AR, illumination change and occlusion [2] in street view videos. Another enhancement for ads insertion is building recognition [5] before ads insertion, from which the contextual ads and user targeted ads come true.

(a) video insertion (b) ads insertion in tennis court

Figure 11. Other examples of ads insertion

5. REFERENCES

[1] B. S. Alvarez, P. Carvalho, M. Gattass, "Insertion of three-dimensional objects in architectural photos", *Journal of WSCG*, 10(1):17-24, 2002.

[2] C. Chang, K. Hsieh, M. Chiang, J. Wu, "Virtual spotlighted advertising for tennis videos", *J. of Visual Communication and Image Representation*, 21(7):595-612, 2010.

[3] X. Li, "Video processing via implicit and mixture motion models", IEEE T-CASVT, 17(8): 953-963, Aug., 2007.

[4] R Hartley and A Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, 2003.

[5] Y. Kim, K. Lee, K. Choi, and S. Cho, "Building Recognition for Augmented Reality Based Navigation System", IEEE Int. Conf. Computer and Information Technology, 2006.

[6] H. Liu, S. Jiang, Q. Huang, C. Xu, "A generic virtual content insertion system based on visual attention analysis", ACM MM'08, pp.379-388, 2008.

[7] G. Medioni, G. Guy, H. Rom, "Real-Time Billboard Substitution in a Video Stream", Digital Communications, 1998.

[8] H. Shah, S. Chaudhuri, "Automated billboard insertion in video", ACCV'07, pp.240-250, 2007.

[9] J. Shi and C. Tomasi. "Good Features to Track". IEEE CVPR'94, pp. 593-600, 1994.

[10] J. Tardif, "Non-Iterative approach for fast and accurate vanishing point detection", IEEE ICCV, pp.1250-1257, 2009.

[11] K. Wan, X. Yan, X. Yu, C. Xu, "Robust goal-mouth detection for virtual content insertion", ACM MM'03, pp.468-469, 2003.

[12] W. Zhang, J. Kosecka, "Extraction, matching and pose recovery based on dominant rectangular structures", IEEE ICCV'03, pp. pp.83-91, 2003.