

A New Natural Marker Detection Method in AR Using Sketch-Based Image Matching

Jeajoon Seho and Junchul Chun ⁺

Department of Computer Science, Kyonggi University, South Korea

Abstract. In this paper, we present a new sketch-based image matching approach to localize a natural marker of an Augmented Reality system. Conventionally, sketch-based image matching is known as a content-based retrieval method to compare the database images with a sketch-based image drawn by users and estimate the degree of similarity between the database images and the query image. In this paper, we adopt sketch-based matching method to localize the natural marker of the sequential images to augment 3D virtual object in AR system. Most similar object in the input image is determined as a natural marker of the AR by comparing the user defined sketched image based on the features such as their edge or corners etc. Unlike other image matching methods, this matching technique is possible to produce query image without constraints by drawing the image intuitively. In addition, in the proposed sketched based AR system, the 3D object augmented on the marker will be also determined by object matching between the detected marker and 3D database images.

Keywords: sketch-based image matching, SURF, feature matching, augmented reality, natural marker.

1. Introduction

Augmented Reality is known to allow virtual graphic models to exactly overlay physical objects in real time. Recently developed tracking and interaction methods in AR allow users to work with and examine the real physical world, while controlling augmented objects in the system conveniently. In general Augmented Reality (AR) can be classified into two categories: marker-based AR and marker-less AR [1], [2], [3] In marker-based AR a specific marker is used for overlaying an object in the scene. Meanwhile, marker-less AR does not require the forethought of adding markers to a scene in order to render a virtual object. It uses a detected feature such as human hand from the scene as a marker instead of using a predetermined specific marker.

The limitation of an artificial marker used in conventional marker-based AR hinders from the generation of the natural scene when the 3D object is overlaid on the marker because the marker itself is awkwardly presented in the image. Thus, the natural marker is preferred to augment the 3D object in AR to obtain more natural output images. In order to use the natural marker, the features of the predefined marker should be collected and analysed in advanced for natural marker detection in augmenting the 3D object. In the view point of the flexibility to select marker in a real-time fashion or user's convenient, the use traditional natural marker has limitation.

Therefore, in this paper we introduce a new sketch-based object matching to localize the natural marker in AR by considering user's convenience to select natural markers in a real-time fashion. Sketch-based image matching, which is known as a content-based retrieval method to compare the database images with sketch images drawn by users, is used to detect a desired object in an input image and the detected object is used as natural maker of AR for augmenting a virtual 3D object.

⁺ Corresponding author. Tel.: + 82-31-249-8947; fax: +82-31-249-8949.
E-mail address: jcchun@kgu.ac.kr.

2. The Proposed Approach

2.1. Description of the Overall Process

The main idea for suggested sketch matching method for Augmented Reality is to obtain a natural marker for AR by the concept of sketch-based image retrieval. It is realized in the following order: getting input video images, sketch matching, object matching and registering a 3D model in AR. Fig. 1 shows the overall procedures of the proposed method.

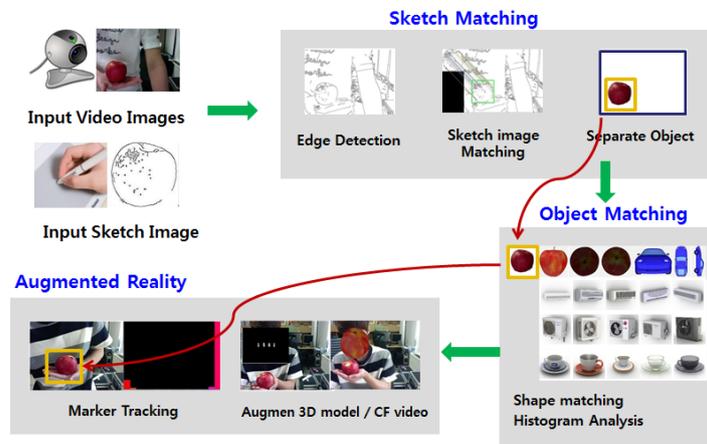


Fig. 1: Overall process for the proposed approach.

In order to detect the natural marker based on the sketch-based matching, the proposed method is classified into three major steps: sketch matching between the sketch image drawn by a user and an object in the input video image, shape matching between detected object and database image and augmentation of virtual object. In the process of sketch matching, we use the SURF algorithm [4], [5] to detect a desired object from input video image. It is known that SURF is a local feature detector and descriptor that can be used for tasks such as object recognition or registration or 3D reconstruction.

In the phase of object matching [6], the overlaid 3D object is dispatched from the database which includes 3D models by comparing the detected marker object with stored objects in database. The 3D database consists of 2D images of the front, side and top view of the various 3D objects to be augmented in AR system. After the exact object are separated from the background through GrabCut [7], in the comparison stage, once each view of 2D images is converted into sketch images and shape and texture matching are processed to measure the similarity between the detected marker and sketch images. In shape matching Hu-moment which is known to be size and rotation invariant is used. In addition, LBP (Local Binary Pattern) [8] of the marker and sketch images is substantially investigated for similarity measure.

While registering a virtual object in the localized natural marker, the detected marker is traced by mean-shift [9] in the video images since the scenes are varying in time. In this work we consider augmented objects as typical 3D model sketched by the user or video images.

2.2. Localization of a Natural Marker

Sketch-based matching is a way to compare features between the input sketch image made by users and the input video images. In order to convert the video image into a sketch, we use canny edge detection algorithm. Fig. 2 shows the results of canny edge detection from two different video images.



Fig. 2: Results of canny edge detection.

When the edge image is obtained, the next step is comparing features between the sketched images by SURF(Speeded Up Robust Features) algorithm. SURF uses an integer approximation of the determinant of Hessian blob detector, which can be computed with 3 integer operations using a precomputed integral image. Its feature descriptor is based on the sum of the Haar wavelet response around the point of interest. In the use of SURF, square-shaped filters are generally used as an approximation of Gaussian smoothing. Integral image is used with a square and defined as:

$$S(x, y) = \sum_{i=0}^x \sum_{j=0}^y I(i, j) \quad (1)$$

The sum of the original image within a rectangle can be evaluated conveniently using the integral image. SURF uses a blob detector based on the Hessian matrix to find points of interest and it also uses the determinant of the Hessian for selecting the scale. Given a point $P(x, y)$ in an image I , the Hessian matrix $H(p, \sigma)$ at point p and scale σ , is defined as follows:

$$H(p, \sigma) = \begin{pmatrix} L_{xx}(p, \sigma) & L_{xy}(p, \sigma) \\ L_{xy}(p, \sigma) & L_{yy}(p, \sigma) \end{pmatrix} \quad (2)$$

where $L_{xx}(p, \sigma)$ etc. are the second-order derivatives of the grayscale image.

Finally the exact shape of the marker can be extracted by GrabCut which utilizes a user-specified bounding box around the object to be segmented. GrabCut estimates the color distribution of the target object and that of the background using a Gaussian mixture model. Fig. 3 illustrates results of SURF and Grabcut. The detected object will be considered as a natural marker of the AR system.

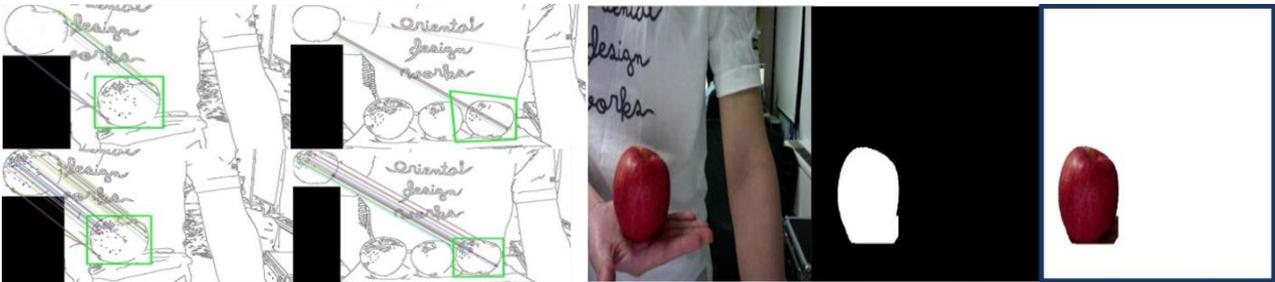


Fig. 3: Object detected by SURF (left) and GrabCut (right).

2.3. Object Matching for AR

In this work, object matching can be achieved by both shape matching and texture analysis of the localized marker of the input video images and the 2D image views of 3D object in the 3D database. The 3D database contains the information of the front, side, top view images of various 3D objects to be augmented in AR system. In shape matching the shape similarity between the boundaries of the marker and 2D images of the 3D object is evaluated based on Hu-moment which includes M_1 to M_6 as absolute orthogonal invariants (independent of position, size, and orientation) and M_7 as a skew orthogonal invariant. These features are utilized to recognize objects in this work.

As a result of the shape matching, some candidate 3D objects are determined from the all objects. Subsequently, the LBP (local binary pattern) operator is used to measure the textural similarity between the marker and 2D images of the candidate objects. These multi-step processes improve the overall matching accuracy. The LBP operator transforms an image into an array or image of integer labels describing small-scale appearance of the image. These labels or their statistics such as the histogram are then used for further image analysis. In general LBP code is determined by comparing the centre pixel value with neighbour pixels of 3×3 image block as follows

$$s(g_0, g_i) = \begin{cases} 1, & g_i \geq g_0 \\ 0, & g_i < g_0 \end{cases} \quad \text{for } i = 1..8 \quad (3)$$

The generic local binary pattern operator is derived from this joint distribution. It is obtained by summing the threshold differences weighted by powers of two as following formula

$$LBP(g_0) = \sum_{i=0}^8 s(g_0, g_i) \cdot 2^{i-1} \quad (4)$$

The best matched 3D object will be registered on the detected marker of the Augmented Reality system. In input video image, the location of the detected marker is varying constantly, thus it is necessary track the location of the marker. In this work, the marker tracking is possible by using mean-shift algorithm[9]. It is an iterative method which starts with an initial estimate x . When a kernel function $K(x_i - x)$ be given, this function determines the weight of nearby points for re-estimating the mean. It estimates the probability density of each class with weighted Gaussian kernel placed on each sample. Each weight is set to be the pixel's posterior probability $P(c_i|I, x)$ corresponding to the class c_i , times the square of the depth of the pixel, which is an estimate of the region the pixel covers. Typical Gaussian kernel on the distance to the current estimate is $K(x_i - x) = e^{-c\|x_i - x\|^2}$. The weighted mean of the density in the window determined by K can be expressed by

$$m(x) = \frac{\sum_{x_i \in N(x)} K(x_i - x)x_i}{\sum_{x_i \in N(x)} K(x_i - x)} \quad (5)$$

where $N(x)$ is the neighborhood of x and $K(x) \neq 0$. The mean shift sets $x \leftarrow m(x)$, and repeats the estimation until $m(x)$ converges. Following Fig. 4 illustrates the procedure of object matching to determine 3D object for AR system.

3. Experimental Results

Fig 4 illustrates the procedure of object matching to determine a 3D object used for AR system.

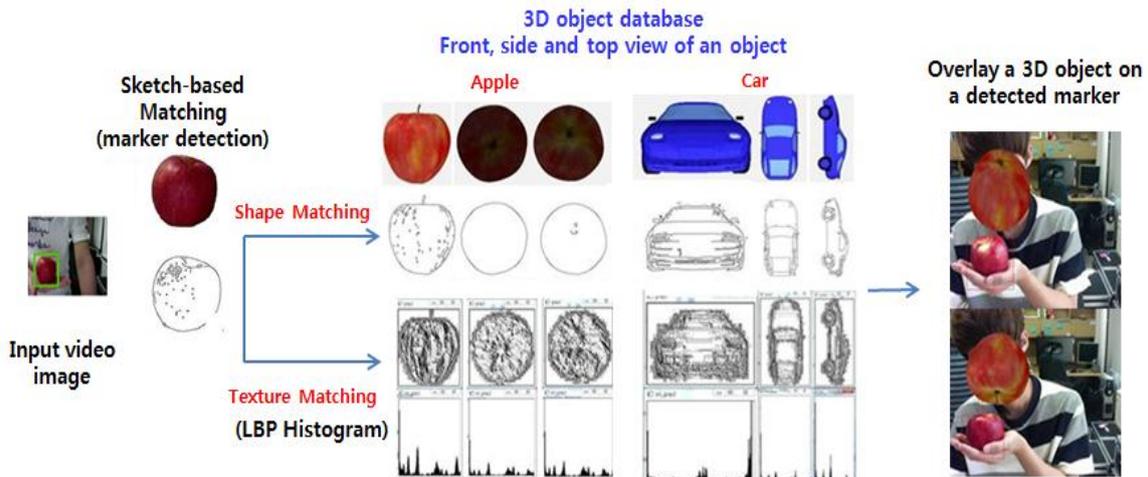


Fig. 4: Object matching process to retrieve 3D model for AR system.

Fig. 5 shows a few samples of 3D models used in database and their corresponding experimental results of sketch images for shape matching, LBP and histograms of LBP for texture matching to retrieve the desired 3D object to be augmented in AR system.

Fig. 6 is shows the final results of augmented objects in AR system with the proposed approach. Both an ellipsoid and an apple are localized in the input image and their corresponding models retrieved from DB are augmented in AR system.

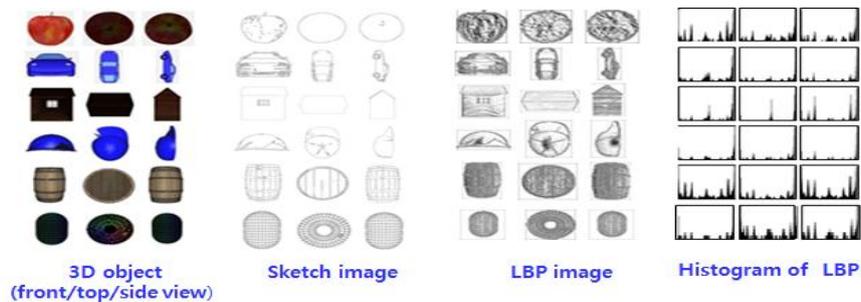


Fig. 5: Samples of 3D object and their experimental results for object matching.



Fig. 6: Results of registering 3D objects on each detected natural marker.

4. Conclusion

This paper proposed a new approach to localize a natural marker of AR by using sketch-based image matching. By using the proposed sketch-based image matching method users can provide query image to localize a natural marker of the AR system from input video images. The augmented 3D object is also determined by evaluating similarity between the stored images of 3D model in database and the detected natural marker based on the object shape and texture matching methods. The results of the work will help to locate the marker of AR and register 3D model in AR system conveniently. As for the future work, how to construct an 3D model augmented using the detected marker itself without the 3D database is major concern.

5. Acknowledgements

This research was supported by GRRC (Gyeonggi Regional Research Centre). The research title is “Development and Industrializing Core-Technologies for Social Services Convergence Platform”.

6. References

- [1] Lee, BS and Chun, JC, Interactive manipulation of augmented objects in marker-less AR using vision-based hand mouse, In Int’l Conference on Information Technology(ITNG), 2010, pp. 398-403
- [2] Chun, J.C and Lee, B.S, Dynamic Manipulation of a Virtual Object in Marker-less AR system Based on Both Human Hands, *KSII Transactions on Internet and Information Systems*, 2010, **4**(4), pp. 618-632.
- [3] JC Chun, SH Lee, A Vision-based 3D Hand Interaction for marker-based AR, *International Journal of Multimedia and Ubiquitous Eng.*, 2012, **7**(3), pp. 51-57.
- [4] H. Bay, T. Tuytelaars, and L.V. Gool. Surf: Speeded up robust features, *ECCV* 2006, pp. 404–417.
- [5] Ahyun Lee et al. Markerless Augmented Reality System based on Planar Object Tracking with SURF, *HCI* 2011, pp. 318-320.
- [6] Al Kabary, Ihab, and Heiko Schuldt. SKETCHify–An Adaptive Prominent Edge Detection Algorithm for Optimized Query-by-Sketch Image Retrieval. *Adaptive Multimedia Retrieval: Semantics, Context, and Adaptation, AMR 2014*, pp. 231-247.
- [7] C. Rother, V. Kolmogorov, and A. Blake. GrabCut: Interactive foreground extraction using iterated graph cuts, *ACM Transactions on Graphics*, 2004. **23**, pp. 309-314.
- [8] Timo Ahonen. Face Description with Local Binary Patterns: Application to Face Recognition, *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 2006 **28**(12), pp. 2037-2041.
- [9] Cheng, Yizong. Mean Shift, Mode Seeking, and Clustering, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1995 **17**(8), pp.790–799.