A Framework of Video News System Using Image Segmentation and Augmented Reality

Jia-Hong Lee, Mei-Yi Wu and Tzu-Hao Tseng

Abstract—**In this paper, we propose a framework of video news system using image segmentation and augmented reality scheme. The goal of proposed system is to provide users with realistic audio-visual contents when they are interested in reading some topics on newspapers. In our approach, pictures on the newspapers were used as markers of augmented reality. Users can pick up these pictures to watch the corresponding video news to obtain more information. The proposed system consists of image segmentation, image recognition and augmented reality engine with audio-visual contents. In image segmentation, skew angle detection and Hough transform techniques are applied to extract pictures on newspapers. In image recognition, image calibration using four-point mapping and Harris corner detection are proposed to identify different pictures on the newspapers. In augmented reality, four-point transformation is reused to add video image frame on the captured image by camera. We expect that the proposed system is popular and can be applied as a kind of advertisement of products in business applications.**

Index Terms— **augmented reality, Hough Transform, document image, image segmentation.**

I. Introduction

MANY studies were proposed into developing Augmented Reality technologies for entertainment, education and advertisement. These applications allow the user to view and manipulate images or virtual 3D objects in a real-world environment. There are two major approaches in camera tracking augmented reality, marker and marker-less methods. Camera tracking systems based on placing markers in the scene have been highly successful[1][2]. Markers are constructed so that they are easily detected in each image frame and given some a priori information about the shapes or positions of the markers, the relative pose of the camera can be easily determined. However, these markers traditionally consist of a set of small white and black squares and are not good looking. In addition, camera tracking can be easily lost as it is only based

Jia-Hong Lee. Author is with Information Management Department, the National Kaohsiung First University of Science and Technology, 2 Jhuoyue Rd., Nanzih, Kaohsiung City, 811, Taiwan (corresponding author, phone: 886-7-6011000; fax: 886-7-6011069; e-mail: jhlee@ nkfust.edu.tw).

Mei-Yi Wu. Author is with General Education Center, National Kaohsiung University of Hospitality and Tourism. No.1, Songhe Rd., Xiaogang Dist., Kaohsiung City 81271, Taiwan (e-mail:barbara @mail.nkuht.edu.tw)

Tzu-Hao Tseng. Author is with Information Management Department, the National Kaohsiung First University of Science and Technology.
.

on a few features and there is a limited range of camera viewpoints from which the markers are visible. In comparison, systems based on natural features, e.g., corner points in the scene extend the tracking range and are typically more stable as there are more features available to track the camera pose from. But it costs much time in computing the large features.

Fig. 1. An example of animated newspapers in Harry Potter films.

In this work, we have developed a framework of video news system based on augmented reality techniques. The system can play video films like the animated newspapers in Harry Potter films. Figure 1 shows the example in Harry Potter films. The pictures on newspapers were regarded as augmented reality markers and we can continuously add a video image frame using image segmentation and augmented reality techniques. Image segmentation operations include thresholding, distorted angle detection and window Hough transform are applied to figure out the shapes and positions of pictures. We divided the picture with many sub-blocks of the same size and compute the image features in each block to form a feature histogram. The frequencies of corner points appeared in different regions in pictures are used as features to identify different pictures. Then an efficient corner tracking with HMMD descriptor is performed to dynamically figure out the position of the distorted picture and add the video news image frame on the captured image by camera on real time.

The overall structure of the proposed system including skew angle detection of newspapers, corner detection, four-point mapping, image recognition and picture tracking is described in Section 2. The related experiments are shown in Section 3. Section 4 shows the conclusion of the proposed system.

II. OVERVIEW OF PROPOSED VIDEO NEWS SYSTEM
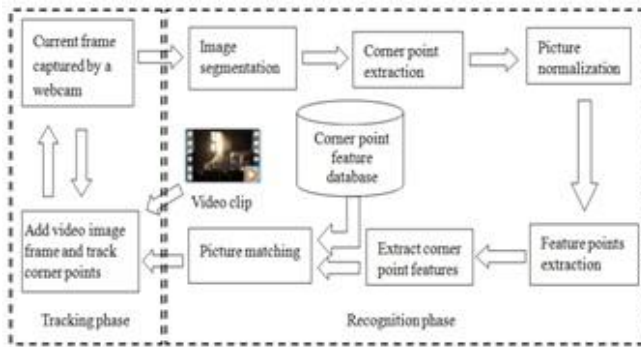
A. Diagram of the proposed video news system

Fig. 2. The block diagram of the proposed video news system.

The proposed system can be divided into two phases: the recognition phase and the tracking phase. Figure 2 shows the block diagram of the proposed system, each captured image frame by the webcam will be applied as input to do an image segmentation procedure. This procedure includes a skew-angle detection and windowed Hough transformation to figure out the boundaries of pictures on a specified newspaper. The four corner points for each extracted picture are determined and a four-point transformation is performed to normalize the extracted picture. Then the picture features are computed according to those extracted feature points. Therefore, we can find the most similar picture in the database and determine the corresponding video news to be played. Once the picture is identified, the system will track the four corner points of this picture to add the corresponding video image frame on it. In the tracking algorithm, a small region R around each corner point is defined and the HMMD color feature is computed for pixels in R. Finally, we transform the video clips by an inverse mapping and put these image frames onto the projected screen of camera to achieve the augmented reality purpose. The image segmentation procedure and other related operations are described in following sections.

B. Skew angle detection

Chou et, al[3] proposed a fast method to detect the skew angle of document images. In their method, a document image was divided into a number of non-overlapping regions, called slabs. One can draw scan lines at all the angles. The lines are further divided into as many sections as the number of slabs, where a section is defined as the part of a scan line that lies within a slab. Then, we can examine each section of these scan lines. If a section S contains at least one black pixel, we change S to gray; otherwise, it stays white. By doing o, we cover all textual and non-textual objects in a slab i^{th} parallelograms that skew at the given angle. Finally, we can count the pixels of white sections of the scan lines drawn at angle θ. The optimal skew angle will be determined if the white area in the obtained image is the largest. Figure 2(a) shows a newspaper image, Figure 2(b) and Figure 2(c) are the result images whenθ=0° and θ=-4°. We can find that the white area in Figure 3(c) is larger than in Figure 3(b), it means thatθ=-4° is a better skew angle than θ=0°.

(a)

(b)

(c)

Fig. 3. Skew newspaper image and the obtained results using Chou's method with different scanning angles.(a) original image (b) result image at angle -3.6 (c) result image at angle 2.

C. Corner detection with windowed Hough transform

In order to detect the corner points of pictures, a fast rectangle detection algorithm using windowed Hough Transform is applied. Most rectangle detection techniques reported in the literature are based on edge and line primitives. Jung et, al[4] proposes a new technique for rectangle detection using a windowed Hough Transform. Every pixel of the image is scanned, and a sliding window is used to compute the Hough Transform of small regions of the image. Peaks of the Hough image are then extracted, and a rectangle is detected when four extracted peaks satisfy certain geometric conditions.

Assume that a rectangle with vertices $P1 = (x1, y1)$, $P2 = (x2, y2)$, $P3 = (x3, y3)$ and $P4 = (x4, y4)$, with $P1P2$ and $P3P4$ being parallel sides with length a, as well as $P2P3$ and $P4P1$

with length b. Also, let us assume that the origin of the coordinate system is located in the center of the rectangle, as shown in Figure 4(a). The corresponding result after Hough Transform is shown in Figure 4(b).
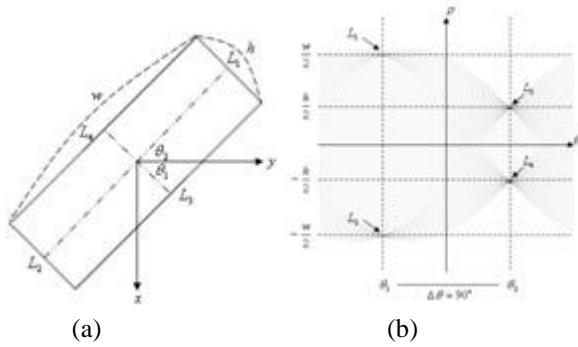

(a)   (b)

Fig. 4. An example of windowed Hough Transform.

It can be observed that these four peaks satisfy the following geometric relations:

1. They appear in pairs: the first one is formed by peaks L1 and L2, at $\theta = \alpha 1$; the second one is formed by peaks L3 and L4, at $\theta = \alpha 0$.
2. Two peaks belonging to the same pair are symmetric with respect to the θ axis, i.e., $\rho 1 + \rho 2 = 0$ and $\rho 3 + \rho 4 = 0$.
3. The two pairs are separated by $\Delta\theta = 90°$ in the θ axis, i.e., $|\alpha 1 - \alpha 0| = 90°$.
4. The heights of two peaks within the same pair are exactly the same, and represent the length of the respective line segment, i.e., $C(\rho 1, \theta 1) = C(\rho 2, \theta 2) = b$ and $C(\rho 3, \theta 3) = C(\rho 4, \theta 4) = a$.
5. The vertical distances (ρ axis) between peaks within each pair are exactly the sides of the rectangle, i.e., $\rho 1 - \rho 2 = w$ and $\rho 3 - \rho 4 = h$.

However, in our application, a picture shape in the captured image frame by the camera will become a distorted rectangle. To solve this situation, we modified the above geometric relations as $|\alpha 1 - \alpha 0| = 90° + T_\theta$, $\rho 1 - \rho 2 = w \pm T_w$ and $\rho 3 - \rho 4 = h \pm T_h$, where T_θ, T_w and T_h are threshold values for distortion tolerance. Finally, we can decide the four corner points from the obtained four lines of a distorted rectangle. Figure 5 is an example of A distorted rectangle and the result after Hough Transform.
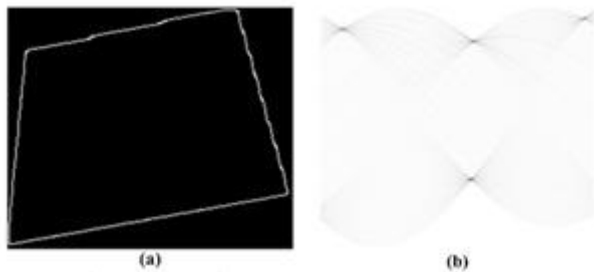

Fig. 5. A distorted rectangle and the result after Hough Transform.

D. Four-point transformation

The quality of the calibration step has a significant influence on the image overlay of video news. In order to achieve the augmented reality goal to show the video news pasted on the captured distorted picture of a camera, a four-point transformation[5] to convert a video image frame to fit the shape of distorted picture is performed using the following equations.

To map from an arbitrary sequence of four 2D points (x_1,y_1), (x_2,y_2), (x_3,y_3),(x_4,y_4) to a set of corresponding points(x_1',y_1'), (x_2',y_2'), (x_3',y_3'), (x_4',y_4'), the transformation requires eight degrees of freedom. The projective transformation can be expressed as a linear mapping in homogeneous coordinates:

$$\begin{pmatrix} x' \\ y' \\ h' \end{pmatrix} = \begin{pmatrix} h'x' \\ h'y' \\ h' \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & 1 \end{pmatrix} \begin{pmatrix} x \\ y \\ 1 \end{pmatrix} \quad (1)$$

We can find the eight unknown transformation parameters $a_{11},\ldots,a_{32}$ by solving a system of linear equation.

$$x_i' = a_{11}x_i + a_{12}y_i + a_{13} - a_{31}x_ix_i' - a_{32}y_ix_i',$$
$$y_i' = a_{21}x_i + a_{22}y_i + a_{23} - a_{31}x_iy_i' - a_{32}y_iy_i', \quad (2)$$

An alternative method for finding the eight parameters for a given set of image points is to use a two-stage mapping through the unit square, which can avoid iteratively solving a system of equations.

$$(0,0) \to x_1' \quad (1,0) \to x_2' \quad (1,1) \to x_3' \quad (0,1) \to x_4'$$
$$x_1' = a_{13}$$
$$y_1' = a_{23}$$
$$x_2' = a_{11} + a_{13} - a_{31} * x_2'$$
$$y_2' = a_{21} + a_{23} - a_{31} * y_2'$$
$$x_3' = a_{11} + a_{12} + a_{13} - a_{31} * x_3' - a_{32} * x_3'$$
$$y_3' = a_{21} + a_{22} + a_{23} - a_{31} * y_3' - a_{32} * y_3'$$
$$x_4' = a_{12} + a_{13} - a_{32} * x_4'$$
$$y_4' = a_{22} + a_{23} - a_{32} * y_4' \quad (3)$$

The set of equations has the following solution for the eight unknown transformation parameters:

$$a_{31} = \frac{(x_1'-x_2'+x_3'-x_4')*(y_4'-y_3')-(y_1'-y_2'+y_3'-y_4')*(x_4'-x_3')}{(x_2'-x_3')*(y_4'-y_3')-(x_4'-x_3')*(y_2'-y_3')}$$
$$a_{22} = \frac{(y_1'-y_2'+y_3'-y_4')*(x_2'-x_3')-(x_1'-x_2'+x_3'-x_4')*(y_2'-y_3')}{(x_2'-x_3')*(y_4'-y_3')-(x_4'-x_3')*(y_2'-y_3')}$$
$$a_{11} = x_2' - x_1' + a_{31} * x_2'$$
$$a_{12} = x_4' - x_1' + a_{32} * x_4'$$
$$a_{13} = x_1'$$
$$a_{21} = y_2' - y_1' + a_{31} * y_2'$$
$$a_{22} = y_4' - y_1' + a_{32} * y_4'$$
$$a_{23} = y_1' \quad (4)$$

Figure 6 shows an example to convert a distorted image into a square shape image using the above four-point mapping.

(a)   (b)

Fig. 6. An example of four-point transformation. (a) the image taken by a webcam (b) the calibrated image using the four-point transform.

E. Image recognition and corner tracking

The Harris corner detector[6] is a popular interest point detector due to its strong invariance to rotation, scale, illumination variation and image noise. In this proposed system, Harris corner features are applied to identify different pictures on newspapers. We divide the calibrated image into M sub-blocks with the same size and count the corner points obtained by Harris corner detection algorithm for each sub-block. These values of corner counts in each sub-block consist of a corner feature histogram H. Let A, B be two images and their corresponding corner feature histograms are H_A and H_b. The distance of two images can be defined as

$$Distance(A, B) = \sum_{i=1}^{M} \frac{|H_A(i) - H_B(i)|}{Max(H_A(i), H_B(i))} \tag{5}$$

The distance value will be small if image A and B are similar.

In contrast, if A and B are not similar, the distance value will be large. We can recognize different pictures in newspapers using the distance measure.

Once the picture is identified, the system will track the four corner points of this picture to add the corresponding video image frame on it. In the tracking algorithm, a small region R around each corner point is defined and the HMMD color feature is computed for pixels in R. HMMD color model (MPEG-7-compatible) is a descriptor for quantized colors, which is claimed to be more uniform than the HSV color space, was proposed by Kim et al. [7]. The HMMD color model was developed from the RGB and the HSV color spaces. We use 128 bins color histogram as the feature for tracking the corner point in the next frame. In the next frame tracking process, we move a window around the corner point with the same size of R and compute the corresponding color histogram of pixels in R. The minimum distance measure of two histograms is used to decide the new locations of the four corners in the next frame. Once the new four corner points are determined, an inverse four-point transformation is applied to an image frame of video news. This transformation can convert a rectangle shape of image frame into a quadrilateral shape determined by the coordinates of the four corner points. Figure 7 shows an example of corner point tracking.

(a)   (b)

Fig. 7.   An example of corner point tracking. (a)front image frame and a blue square R centered at a corner point (b)current image frame and a red square R' centered at the corner point was detected using mean shift tracking algorithm..

III.   EXPERIMENTAL RESULTS AND DISCUSSIONS
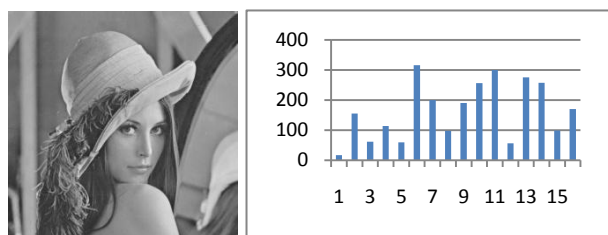
A. Image segmentation

To evaluate the proposed image segmentation method, different newspapers are applied for image segmentation. Figure 8 shows some examples and the pictures in newspapers can be segmented very completely.
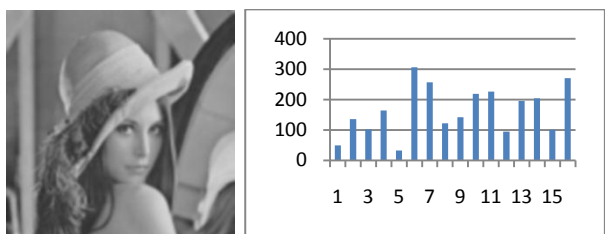

(a)

(b)

Fig. 8. Picture extraction results surrounded with red lines using the proposed image segmentation method. (a) picture extraction from two images captured by a webcam in different views; (b) picture extraction from four different newspapers captured by a webcam.
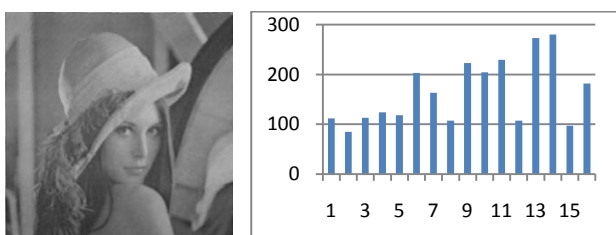
B. Image recognition

To evaluate the performance of proposed features using Harris corner histograms, six images of size 512 by 512 are used for tests. There are one of ten pixels (2621 pixels) were marked as Harris corner points and the feature histogram to count the number of Harris corners located in 16 different regions is derived. Figure 8 shows the images and the computed Harris corner histograms. It is clear that the histograms of Lena image, its blur image and the calibrated image are very similar. The distances between different images were listed in TABLE I. It shows that Harris corner histogram can achieve a good performance for identifying different images.
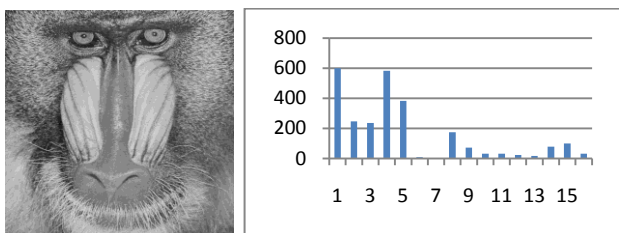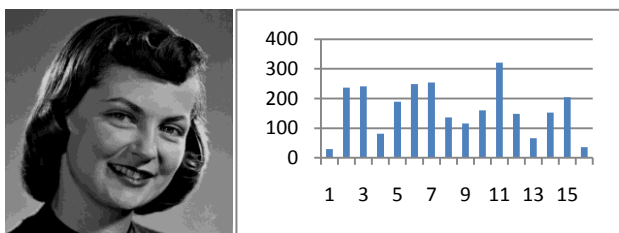
(A) Original image

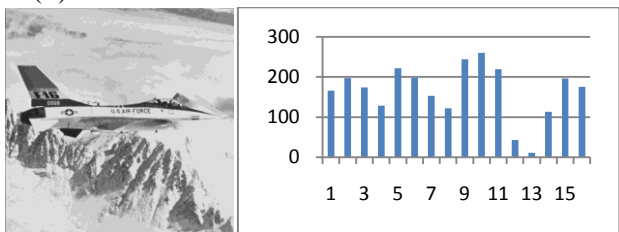
(B) Blur image by Gaussian filter (σ=3)

(C) Calibrated image by four-point mapping

(D) Baboon

(E) Girl

(F) Jet

Fig. 8. Test images and their corresponding Harris corner point histograms using the proposed method.

TABLE I
THE DISTANCES BETWEEN DIFFERENT IMAGES

	A	B	C	D	E	F
A	0	**4.34**	**4.59**	11.56	7.12	6.20
B	**4.34**	0	**1.37**	11.20	6.42	6.64
C	**4.59**	**1.37**	0	11.21	6.07	6.92
D	11.56	11.20	11.21	0	9.29	9.53
E	7.12	6.42	6.07	9.29	0	6.35
F	6.20	6.64	6.92	9.53	6.35	0

The histogram of Baboon seems to be very different to other histograms. Figure 9 shows the locations of Harris corner points which are divided into 16 blocks. The distribution of corners are not uniform, most corner points appeared at the upside of image. If we order these blocks in horizontal scan from up to down, most points are located in the first and fourth blocks and just few points are located in the fifth and sixth blocks.
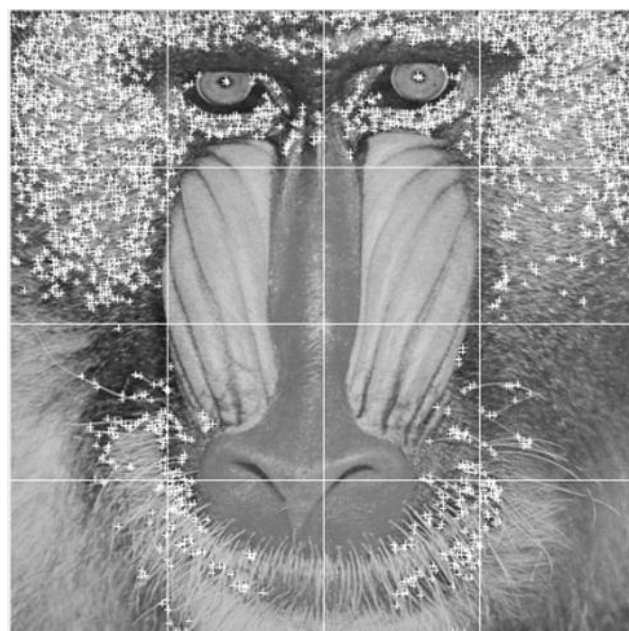

Fig. 9. The locations of Harris corner points for image Baboon.

C. Augmented reality video news

Once the picture is identified, the system will track the four corner points of this picture to add the corresponding video image frame on it. Figure 10 is an example to display the AR video news using the proposed system. Figure 10(b) shows the picture extraction using the proposed image segmentation method. Figure 10(c) is the calibrated image converted from the distorted image of Figure 10(b) using four-point mapping. Figure 10(d) is the Harris corner point histogram of the gray-level image in Figure 10(c). Figure 10(e) is the video image frame and the image is converted into distorted form after inverse four-point mapping and add to the original captured picture. The result of augmented reality is shown in Figure 10(f).
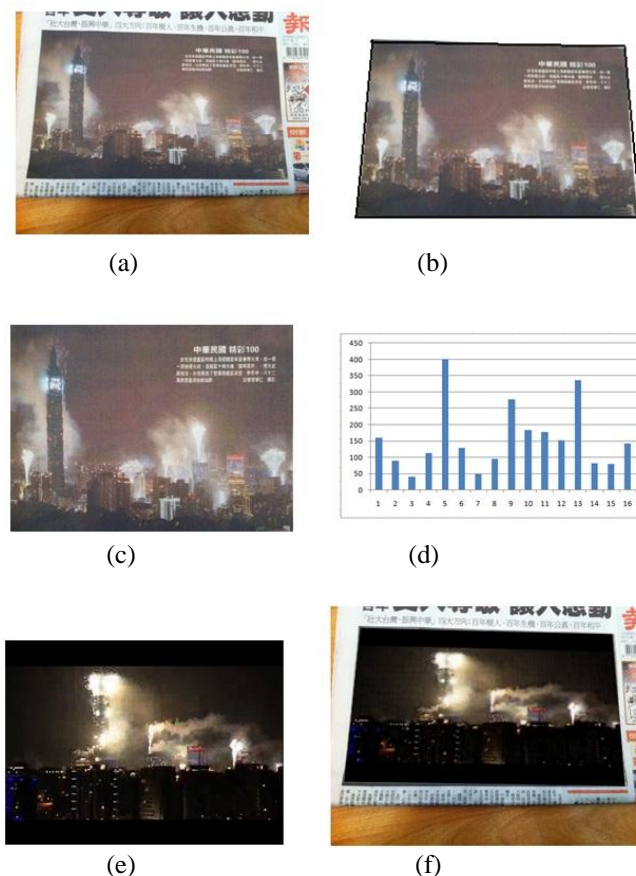
(a)

(b)

(c)

(d)

(e)

(f)

Fig. 10. An example of the proposed AR video news system. (a) the newspaper image; (b) picture extraction using image segmentation; (c) calibrated image of (b) using four-point mapping; (d) the corresponding Harris corner point histogram of the image in (c); (e) the video new image frame;(f) the final result of the system by adding the video image frame of (e) onto the picture of (a),

IV. CONCLUSION

We have proposed a framework of video news system using image segmentation and augmented reality scheme. In our approach, pictures on the newspapers were used as markers of augmented reality. Users can pick up these pictures to watch the corresponding video news to obtain more information. The proposed system consists of image segmentation, image recognition and image calibration. We presented a new image feature using four-point mapping for image calibration and corner point histogram to identify different pictures and performed good performance. We expect that the proposed system is popular and can be applied as a kind of advertisement of products in business applications.

REFERENCES

[1] K. Xu, K. W. Chia, A. D. Cheok,"Real-time camera tracking for marker-less and unprepared augmented reality environments", Image and Vision Computing, Vol. 26, pp. 673-689, 2008.
[2] C. Celozzi, G. Paravati, A. Sanna, and F. Lamberti,"A 6-DOF ARTag-Based Tracking System" , IEEE Transactions on Consumer Electronics, Vol. 56, No. 1, pp. 203-210, 2010
[3] C. H. Chou, S. Y. Chu, and F. Chang, "Estimation of skew angles for scanned documents based on piecewise covering by parallelograms", Patten Recognition, Vol. 40, pp443-455, 2007.
[4] C. R. Jung and R. Schramm,"Rectangle Detection based on a Windowed Hough Transform", Proceedings of the XVII Brazilian Symposium on Computer Graphics and Image Processing (SIBGRAPI'04), 2004.
[5] W. Burger, M.J. Burge, Digital Image Processing-An Algorithm Introduction Using Java, Springer Science+Business Media,LLC, 2008.
[6] C. Harris and M.J. Stephens. A combined corner and edge detector. Alvey Vision Conference, pp. 147–152, 1988.
[7] H. Kim, J. S. Lee, S. Jun, J. M. Song, H. Y. Lee, "Descriptorfor quantized color using HMMD color model,"ISO/IEC/JTCI/SC291 WG II, Lancaster, UK, Feb. pp. 669, 1999.