KEYPOINT-BASED NEAR-DUPLICATE IMAGES DETECTION USING AFFINE INVARIANT FEATURE AND COLOR MATCHING

Yue Wang, ZuJun Hou, and Karianto Leman

Institute for Infocomm Research
A*Star (Agency for Science, Technology and Research), 1 Fusionopolis Way, Singapore
{ywang, zhou, karianto}@i2r.a-star.edu.sg

ABSTRACT

This paper presents a new keypoint-based approach to near-duplicate images detection. It consists of three steps. Firstly, the keypoints of images are extracted and then matched. Secondly, the matched keypoints are voted for estimation of affine transform based on an affine invariant ratio of normalized lengths. Finally, to further confirm the matching, the color histograms of areas formed by matched keypoints in two images are compared. This method has the advantage for handling the case when there are only a few matched keypoints. The proposed algorithm has been tested on Columbia dataset and conducted the quantitative comparison with RANdom SAmple Consensus (RANSAC) algorithm and Scale-Rotation Invariant Pattern Entropy (SR-PE) algorithm. The experiment result turns out that the proposed method compares favorably against the state-of-the-arts.

Index Terms**—**Near-duplicate detection, image matching, color matching, affine invariant feature.

1. INTRODUCTION

Near-Duplicate (ND) image detection aims to evaluate a pair of images which one is close or partially close to the other. It has been widely applied to image and video content analysis, such as location recognition, image spam detection and illegal copy of images and videos detection. A number of techniques have been proposed and can roughly be classified into two categories [6]:

1) Global method (appearance-based method) [1-5]: such as color moments and color histogram from whole image. This kind of techniques is very efficient for finding identical copies, but may be sensitive to the variation of lighting and viewpoint, or occlusions.
2) Local method (keypoint-based method) [6-9]: the method usually detects local keypoints in two images and measures their similarity by counting the number of correct correspondences between two sets of keypoints. This method can deal with the illumination variation and geometric transformation but at the expense of computational efficiency.

Mikolajczyk and Schmid have conducted a comparison of several local descriptors [10]. For a brief account of the latest development on ND image detection, interested readers can refer to [5] or [6].

This paper will present a method to combine the advantages of appearance-based method and keypoint-based method for affine ND image detection. This method is advantageous in handling the case when there are only a few matched keypoints. The remaining structure of this paper is arranged as follows. Our method is detailed in Section 2. Experimental results and discussion are presented in Section 3. Finally, the conclusion of this paper is presented in Section 4.

2. NEAR-DUPLICATE DETECTION USING AFFINE INVARIANT FEATURE AND COLOR HISTOGRAM

Normally there are many keypoint matching lines for two duplicate images which are only slightly changes in rotation and scale, see Fig. 1(a). In this case, a simple threshold for the number of matching lines is able to achieve a good performance. However, when two images have significant change in rotation, scale or illumination, the number of matching lines is small (e.g. Fig. 1(b)). Here we present a method which is able to handle the case of the less matching lines while remaining accurate ND detection. It is based on preliminary filtering with an affine invariant feature and followed by color histogram confirmation. We assume two ND images are related through an affine transformation with certain illumination variation.

2.1. Matching lines filter based on affine invariant feature

In general, an affine transformation consists of a linear transformation (rotation, scaling or shear) followed by a translation:

$$\boldsymbol{T}(\mathbf{x}): \mathbf{x} \mapsto \mathbf{A}\mathbf{x} + \mathbf{b}. \quad (1)$$

There are some properties for two images if they are related with an affine transformation in Euclidean space, which include the preservation of the collinearity relation between points as well as the ratio of distances for distinct collinear points. Parallel lines will remain parallel. Most interestingly, the ratio of the area will be a constant. This property has been utilized in [19] for image matching, where the ratio of areas are formed by three matched pairs.

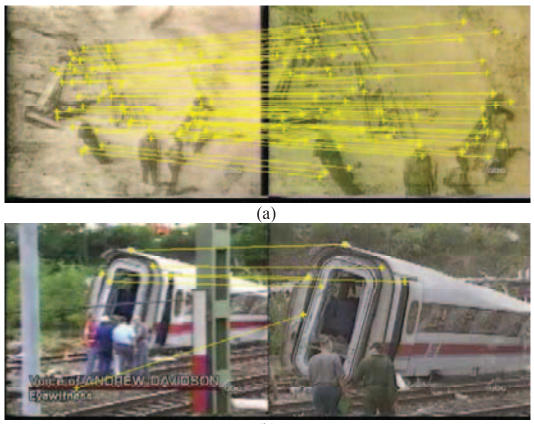


Fig. 1. Keypoints matching. (a) Many matching lines; (b) Fewer matching lines.

In this study, a ratio, which is using the altitude/height of triangles, is employed as the feature for matching, which remains invariant under affine transformation. In addition, compared with the ratio of area, the ratio of distance is able to reduce the computation load in voting. The details are described as follows.

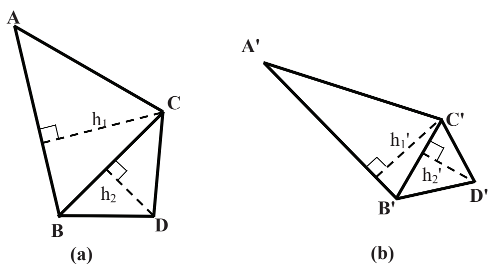


Fig. 2. Affine transformed triangle.

In Fig. 2, the quadrilaterals $ABCD$ and $A'B'C'D'$ are related with an affine transformation: $\boldsymbol{T}(ABCD) = A'B'C'D'$, where h_1 and h_2 are the heights of triangle ABC and BCD, and the heights of the transformed triangles are h_1' and h_2' respectively. Then, by the property of affine transformation, we have

$$\frac{Area_{ABC}}{Area_{A'B'C'}} = \frac{Area_{BCD}}{Area_{B'C'D'}} = \alpha \tag{2}$$

where α is a constant. Eq. (1) can be written as

$$\frac{AB \times h_1}{A'B' \times h_1'} = \frac{BC \times h_2}{B'C' \times h_2'} = \alpha \tag{3}$$

If we rescale h_1' and h_2' as

$$\hat{h}_1' = \frac{A'B' \times h_1}{AB}, \qquad \hat{h}_2' = \frac{B'C' \times h_2}{BC} \tag{4}$$

then we have

$$\frac{h_1}{\hat{h}_1'} = \frac{h_2}{\hat{h}_2'} = \alpha \tag{5}$$

As aforementioned, the affine transformation preserves the ratio α, thus, those heights of triangles under the same affine transformation will lead to the same ratio, indicating the similarity among the key points associated with these triangles. On the other hand, those heights of triangles under different affine transformation will yield different ratios, which suggest that these key points are less correlated. If we divide the ratio into some number of categories (bins) and count the occurrence frequency of each category, we will arrive at a histogram as shown in Fig. 3. In this graphical display, areas under the same affine transformation will vote for the same bin and the votes of areas under different affine transformation will scatter among different bins, appearing like background/noise in the histogram, thus we can infer the degree of ND from the analysis of this plot. If a peak is identified with significantly large value, the confidence on the near-duplicate between two images would be high. Otherwise, the two images will be less likely near-duplicate.

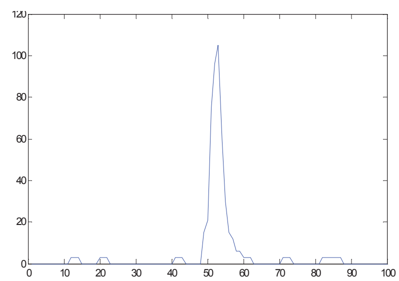


Fig. 3. Histogram of affine transformation for a correct matching image pair, where the horizontal axis is the value of α and the vertical axis is the occurrence frequency.

In our implementation, we design our bin space that can handle the ratio of Eq. (5) up to 5(correspondingly, the ratio of area would be up to 25). The number of bins in the histogram is set to 100, where 50 bins are for the ratio ranging from 1 to 5, and another 50 bins are for the ratio of 0.2 to 1. With this design, it is able to check either all or parts of the matching in two images. It was experimentally determined that detection of a true peak over the noise requires the peak to be at least 10 times greater than the noise level. In our application, the histogram is created with exhaust sampling of any two matched keypoints to form a side of triangle, e.g. AB and $A'B'$ in Fig. 2, and then sample anyone of the rest matched keypoints to form a triangle and calculate the heights of triangle, and get the ratio of Eq. (5).

2.2. Color Histogram Matching

As aforementioned, the number of matched lines could be very small so that the confidence of ND detection based on the number of matched lines would be low. To address this issue, color histogram matching is employed for further investigation. There are various methods for color similarity measurement. Here the normalized color histogram and Bhattacharyya coefficient [14] is employed. Referring to Fig. 4, the region of interest (ROI) as formed by matched keypoints is extracted from two images respectively. Then within the ROI the color of each pixel is normalized in order to attenuate the illumination variation by

$$(R', G', B') = \frac{(R, G, B)}{R + G + B} \quad (6)$$

After that, histogram is generated with 16 bins for each color channel. Then, the matrix of size 16x16x16 is reshaped as a one dimension vector with 4096 elements. For comparing two histograms, (c_i and v_i, i=1 to 4096), the color similarity measurement based on Bhattacharyya coefficient is as follows,

$$S = \sum_{i=1}^{4096} \sqrt{c_i \cdot v_i} \quad (7)$$

The value returned is from 0 to 1. A threshold of 0.6 is utilized in this study, above which will be identified as similar in color.

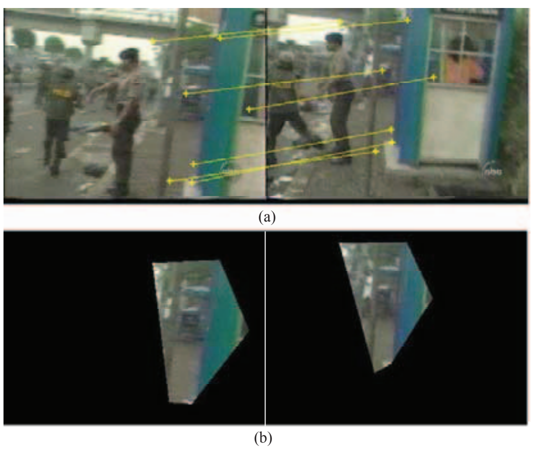


Fig. 4 Color histogram matching. (a) Original images with matching lines. (b) The areas for color histogram matching. The measurement of color similarity for two showed regions is 0.8687.

2.3 Complete Algorithm

The complete algorithm using both affine invariant feature and color similarity measurement for two images is as follows:

(1) Keypoints extraction using SIFT;
(2) Keypoint matching, remove multi-matching points;
(3) If there are more than 30 matching lines, two images are ND and go to step 7.
(4) Conduct the affine transform estimation based on affine invariant feature mentioned in section 2.1. If failed, go to step 7;
(5) If remained matching lines are more than 10, two images are ND and go to step 7.
(6) Perform color histogram matching mentioned in section 2.2;
(7) Stop.

3. EXPERIMENTAL RESULTS AND DISCUSSION

The Columbia dataset, which is selected from TRECVID 2003 video corpus [12], is utilized to validate the effectiveness of the proposed method. It consists of 600 key frames with 179700 candidate pairs. In this dataset, 215 pairs are selected as ND pairs. The groundtruth can be viewed at http://www1.i2r.a-star.edu.sg/~ywang/demo/columbia_groundtruth_215.

In this experiment, the ND detection is achieved by performing an exhaustive search throughout the dataset. We compare the results of our method to RANSAC [14] for affine transform estimation. The Matlab codes for RANSAC is downloaded from [15] and the default settings are used. In addition, we also compare our method with SR-PE [11], which has been reported to have better performance over others. We use the Recall, Precision and F-measure [11] to quantitatively evaluate the performance of ND detection,

$$\text{Recall} = \frac{\text{Number of ND pairs correctly detected}}{\text{Total Number of ND pairs}} \quad (8)$$

$$\text{Precision} = \frac{\text{Number of ND pairs correctly detected}}{\text{Number of detected ND pairs}} \quad (9)$$

$$\text{F - measure} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (10)$$

The effectiveness of the color histogram matching is firstly evaluated. Table 1 shows the results of RANSAC and the proposed method without color matching, where the number of correct ND detected is similar. The proposed method for affine transform estimation is slightly better than RANSAC in terms of lower false detection rate. Table 2 shows the results for both methods where color matching was applied. From there it can be observed that, the false alarms have been remarkably reduced. Table 3 shows the performance comparison of our method with SR-PE on Columbia dataset. It can be observed that although SR-PE is able to achieve a perfect precision, our method is able to get the better recall and F-measure. However, there are three false alarms in our method, as showed in Fig. 5. In Fig. 5(a), the two images are similar in object, background and color, and the matched keypoints are well located on good corresponding places, but they are not the ND in ground truth of Columbia dataset because of different people presented. Fig. 5(b) and (c) are false matching due to few matching lines and similar color tone in the sampled areas.

Table 1. Comparison of the proposed method vs RANSAC (both without color matching).

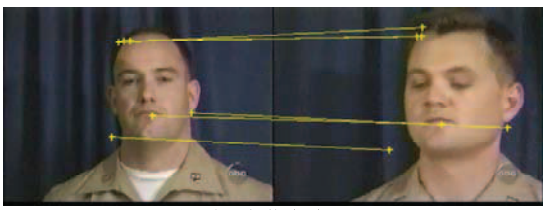
	RANSAC w/o color matching	Proposed method w/o color matching
Pairs detected	242	223
Correctly detected	177	179
False detected	65	44
Precision	0.7314	**0.8027**
Recall	0.8233	**0.8326**
F-measure	0.7746	**0.8174**

Table 2. Comparison of the proposed method vs RANSAC (both with color matching).

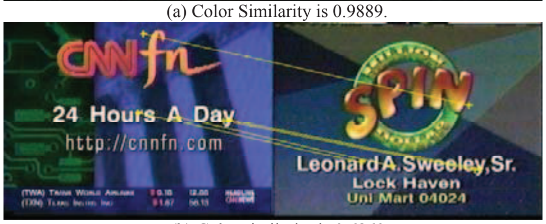
	RANSAC with color matching	Proposed method
Pairs detected	180	180
Correctly detected	169	177
False detected	11	3
Precision	0.9389	**0.9833**
Recall	0.7860	**0.8233**
F-measure	0.8557	**0.8962**

Table 3. Comparison of the proposed method vs SR-PE.

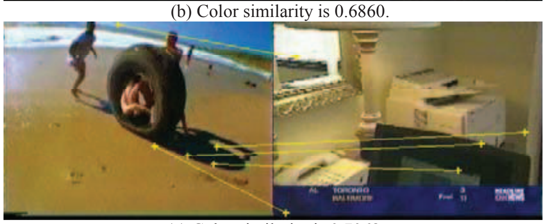
	Proposed method	SR-PE
Pairs detected	180	173
Correctly detected	177	173
False detected	3	0
Precision	0.9833	**1**
Recall	**0.8233**	0.8047
F-measure	**0.8962**	0.8918



(a) Color Similarity is 0.9889.



(b) Color similarity is 0.6860.



(c) Color similarity is 0.7068.

Fig. 5. Three false ND detections using the proposed method.

4. CONCLUSION

This paper presents our works for ND detection. Our algorithm has two main novelties. Firstly, it conducts the estimation of affine invariant based on a ratio of length which is affine invariant and able to reduce the bin space. Secondly, a color histogram matching for the areas formed by the matched keypoints in two images being compared is carried on to ensure the correct matching in appearance. Our method has potential to work with the matching of fewer keypoints, which means we can be more emphasizing on detecting the reliable and salient keypoints, and matching the keypoints with more strong constraint. The qualitative comparison with SR-PE in the experiment shows that the proposed algorithm is on par with the state-of-the-art. Our future work will involve more specific detail to enhance the color matching in order to reduce the false alarm.

5. ACKNOWLEDGEMENT

The authors would like to thank Lim Kian Guan, Edward Chen Jun Cheng, and Kang Shian Chin from the Future System Directorate, MINDEF Singapore, for their contributions in the experiment.

REFERENCES

[1] K. Mikoljczyk, C. Schmid, “Scale and Affine Invariant Interest Point Detectors,” *International Journal of Computer Vision*, vol. 60, pp. 63–86, 2004.

[2] D. G. Lowe, “Distinctive Image Features from Scale-Invariant Keypoints,” *International Journal of Computer Vision*, 60(2):91–110, 2004.

[3] Zhao, W.L., Ngo, C.W., Tan, H.K., Wu, X., “Near-Duplicate Keyframe Identification with Interest Point Matching and Pattern Learning,” *IEEE.Trans. on Multimedia*, vol. 9, no. 5, pp. 1037–1048, 2007.

[4] Ngo, C.W., Zhao, W.L., Jiang, Y.G., “Fast Tracking of Near-Duplicate Keyframes in Broadcast Domain with Transitivity Propagation,” *ACM Multimedia,* pp. 845–854, 2006.

[5] Zhao, W.L., Ngo, C.W., “Scale-Rotation Invariant Pattern Entropy for Keypoint-based Near-Duplicate Detection,” *IEEE Trans on Image Processing*, vol. 18(2), pp. 412-423, 2009.

[6] J. Zhu, Steven C.H.Hoi, Michael R. Lyu and S. Yan, “Near-Duplicate Keyframe Retrieval by Nonrigid Image Matching,” *MM2008*, pp. 41-50, 2008.

[7] A. Qamra, Y. Meng and E. Y. Chang, “Enhanced Perceptual Distance Functions and Indexing for Image Replica Recognition,” *IEEE. Trans. on PAMI* 27(3): 379-391, 2005.

[8] D. Q. Zhang and S. F. Chang, “Detecting Image Near-Duplicate by Stochastic Attributed Relational Graph Matching with Learning,” *ACM MM*, pp. 877-884, 2004.

[9] W. Zhao, Y. Jiang and C. Ngo, “Keyframe Retrieval by Keypoints: Can Point-to-point Matching Help?” *CIVR* 2006, pp. 72-81, 2006.

[10] K. Mikolajczyk, C. Schmid, “A Performance Evaluation of Local Descriptors,” *IEEE Trans. on PAMI*, vol. 27, no. 10, pp. 1615–1630, 2005.

[11] Zhao, W.L., Ngo, C.W., “Scale-Rotation Invariant Pattern Entropy for Keypoint-based Near-Duplicate Detection,” *IEEE Trans on Image Processing*, vol. 18(2), pp. 412-423, 2009.

[12] TREC Video Retrieval Evaluation (TRECVID), http://trecvid.nist.gov/. (2010)

[13] Fischler, M.A., Bolles, R.C., “Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography,” *Commun. ACM* 24(6) 381-395, 1981.

[14] K. Fukunaga, *Introduction to Statistical Pattern Recognition*, Academic Press, second edition, 1990.

[15] http://www.csse.uwa.edu.au/~pk/Research/MatlabFns/index.html.(2010)