LOGO DETECTION IN HIGH-MOTION SPORTS VIDEO

Andre Filgueiras de Araujo, Stephanie Pancoast

{afaraujo, pancoast}@stanford.edu
Stanford University

ABSTRACT

A system to detect logos in the high-motion setting of a sports video is presented, which allows for automated advertisement efficiency verification. We first incorporate a basic feature-matching algorithm using SIFT, nearest-neighbor matching and RANSAC. The main contribution of this work is the capitalization of the temporal redundancy, inherent in videos, by employing a second-pass to propagate features from adjacent frames. The match is performed, in this case, with a more adequate logo to the particular point in the video. Results show that the second-pass improves performance in the detection by capturing on average 20% of the missing frames that result from the first pass.

1. INTRODUCTION

Within the past decade logo detection for sports videos has become an area of increasing interest. The marketing industry invests a large sum of money to place advertisements on billboards, fields, and other items in sporting events. Therefore, feedback on how frequently the logo is actually displayed within the video is valuable.

Consisting of a wide combination of text and shapes, logo detection poses a challenging problem for image processing researchers. Initial studies focused on applications with text documents [1]. Invariants are used to describe the image and affine transformations to refine the match. While this approach is useful, the assumption that the image will appear in a clean, unobstructed setting is not valid in practical settings. For the application of detection in sport video, analysis is performed on individual frames [5, 7, 8, 9]. Authors [9] use a template matching approach on video stills from a soccer game. After determining an area of high contrast, a straight line is processed horizontally across the logo and this is used as that logo's profile for later matching tasks. This method is useful for text-only, clear-shot logos and would not be able to overcome the frequent occlusion and blurriness that is characteristic of a high-motion sports video. In the majority of other works, key-points are first found using the Scale-Invariant Feature Transform (SIFT) [3] on both the frame and a clean version of the logo. A bi-directional matching method is proposed to find a more robust match of features [7]. Matches from the database logo and from the video are looked at to ensure that these matches appear in consistent locations within the logo. Random Sample Consensus (RANSAC) [4] is then used to further eliminate outlying features. In one study [8], preprocessing on the video still is first done to estimate the quality of a frame before continuing to extract and compare SIFT features.

Although much of the previous research has sought to determine the presence of logos in videos, few have actually taken advantage of the temporal redundancy that is inherent in videos. Aside from sporting events, another recent application is for logo detection with hand-held devices to identify the logo and then link the user to deals and information about that company or product [6]. The logo is first identified using the basic SIFT feature extraction and matching, and then continues to track the item using the colors present. This method, while efficient, is not equipped to handle the frequent lighting and perspective changes of a sports video. In a more related work [10], temporal redundancy is used by looking for a logo in frames spaced 10 apart. If a logo is identified in these two frames, it is assumed that those in between also contain the logo. This is useful for shots where the logo is clearly present on a billboard, but would likely result in a number of misses if the majority of the frames were not to contain a clear shot.

The research to date on logo detection in sports videos often relies on clear views of the logo where the type of sporting event is known. However, this is often not the case. The camera is focused on the players and the game, so logos appear in the background, often blurred or occluded. We have explored a system to detect logos in a high-motion sporting event where the majority of frames contain such non-ideal conditions. There are two main steps to the system. We refer to the first pass as the "Basic Algorithm", which uses a similar frame-by-frame technique to previous work [7].The second step is referred to as the "Extended Algorithm". Here, the temporal redundancy of video is used to replace the features from the clean logo with that from the adjacent frames to achieve a better match. In this paper, we first explain the approach in further details, in section 2. The experimental setup to test the system will then be presented in section 3, together with the results.

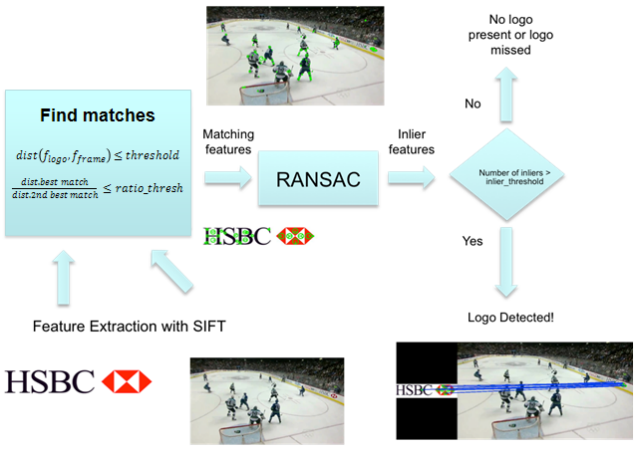

Fig. 1. Block diagram of the Basic Algorithm

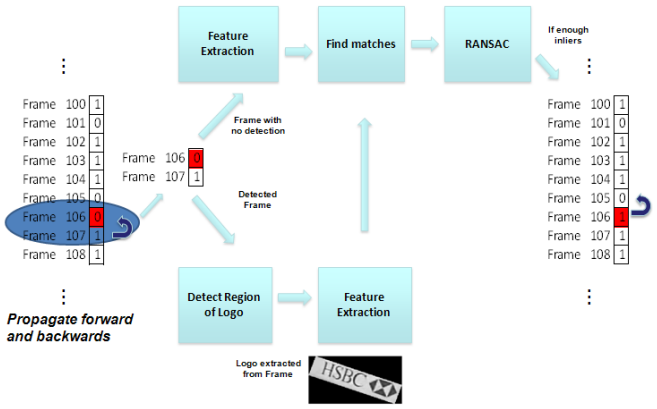

Fig. 2. Illustration of the Extended Algorithm

2. OVERVIEW OF THE SYSTEM

As previously mentioned, the developed system consists of two core parts: "Basic Algorithm" and "Extended Algorithm". The former refers to the well-known image retrieval pipeline using SIFT, and the latter concerns an extension that utilizes the redundant temporal information, inherent to a video source, to improve the performance. We will describe both these stages in the following subsections.

2.1. Basic Algorithm

This stage basically matches each frame to the original logo separately, as illustrated in Figure 1, and is the first pass of the implemented system.

Initially, SIFT features are extracted from both the logo and the query frame. A nearest-neighbour matching between the features from both images is then performed, followed by a ratio test, which will filter the features by their distinctiveness. The set of resulting matched features are passed to a geometric consistency verification task, which is performed by RANSAC. It outputs an affine transformation that links spatially the sets of features from both logos, as well as the number of features that are consistent with it (number of inliers), if any.

All these steps are regulated by thresholds that are crucial for the system to perform reasonably. In general, looser criteria may increase the number of correct matches, but often in conjunction with the detection of false positives. In this work, we adopt thresholds that avoid the detection of false positives.

2.2. Extended Algorithm

The Extended Algorithm is a second pass over the results of the Basic Algorithm, aiming to propagate the matches that were initially detected. It is a first swing at incorporating the inherent temporal redundancy of video that could be used to enhance detection performance.

This approach relies on the fact that it is very likely for the logo to be present in frame $N+1$, if it was present in frame N. Building on the results of the first pass, we basically look for the logo in adjacent frames to the ones for which a match was detected.

When a frame for which a match was declared is found (frame N), we check if the adjacent frame $N-1$ also has a match. In case that it does not, we will try to find the logo in it, given that this is very likely to occur. We already know that there were not enough feature matches between the database logo and frame $N-1$ for a detection to be declared. The key insight that is employed in our approach is, then, the utilization of the logo in frame N as the new template to be searched for, as it will be very similar the logo in frame $N-1$, if present. This is possible due to the results of the first pass: the affine transformation that links the original logo to its version in frame N allows for simple extraction. The steps that are employed in the first pass are then repeated in this case, except that, instead of the original logo, the logo detected from frame N is used. The same process is repeated for $N+1$. Figure 2 illustrates this process for frame 107, which, in the example, has a match, and frame 106, which does not.

Two variants of the Extended Algorithm are implemented. The first one replicates the process for frame $N-1$, after declaring it a match due to the comparison to frame N. That is to say, the logo is extracted from frame $N-1$ and searched for in frame $N-2$, in case frame $N-2$ didn't have a match from the first pass. In the example from Figure 2, the logo would be extracted from logo 106 to be searched for in frame 105.

The second variant of the Extended Algorithm employs only the frame that had a match as a result of the first pass to get a new logo template that will be propagated. In other words, the logo in frame N is used to find matches in frames $N-1$, $N-2$, etc, as long as there is a match for each comparison (and, evidently, as long as it did not have a match from

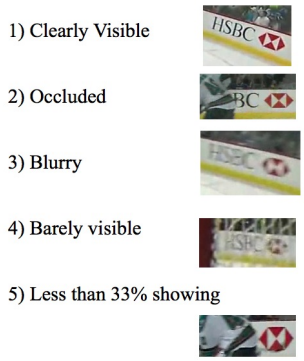

Fig. 3. Categories of the logos present in the video

the Basic Algorithm). In Figure 2, the template used for the search in frame 105 would be the one extracted from frame 107.

As discussed in section 3, there is a trade-off between the two approaches. When the propagation of the logo is performed for every individual frame, the likelihood of finding a false positive increases. This is due to the imperfections that accumulate from each individual logo extraction. As the feature matching and affine model are not perfect, the errors sum up for every new logo extraction, and then regions that are adjacent to the logo can be accidentally captured. From that point on, these regions (e.g., a hockey player) will be considered as part of the logo and, then, will be used as a template for matching.

The second approach has lower probability of detecting false positives because it employs only the logo in the frame that was detected in the first pass. In this case, it is very likely that this logo does not have clutter that would introduce false positives. However, as the matching is held further away from N, the logo to be detected is not very similar to this new template any more. The lower detection of false positives comes at the expense of lower recall performance.

3. EXPERIMENTS

Our experiments consisted of using the implemented algorithm for a hockey game video. The source video was downloaded from *YouTube* at 720p High-Definition resolution, at 30 frames per second, with total duration of 3 minutes and 7 seconds and a total of 5589 frames [11]. The experiments were carried out for the logos HSBC and SONY. The *VLFEAT* [2] *MATLAB*-based implementation of SIFT was employed, and the whole system was implemented and tested under *MATLAB*.

Initially, the video was labeled with five categories representing the visibility of the logos, as shown in Figure 3. Categories 1 to 3 are the most important ones, because a human viewer can clearly identify a logo pertaining to each of those. Then, we consider that a miss occurs if one frame in this category is not found. Categories 4 and 5 represent instances of the logo that can hardly be recognized by a human viewer. Though they are not taken into account for the recall calculation, we do not consider that a false positive occurs if the method declares that a match is found for it. Figure 4 illustrates the detection of the logo in frame 929, with blue lines connecting the pairs of matching features.

Figure 5 and Table 1 summarize the experimental results in terms of recall performance for the given video and both logos. The Basic Algorithm performs poorly, due to many of the reasons that were already mentioned in this report: blurring, frequent perspective changes and occlusion. The logos are not the focus of attention of the sports event and is always on the background. The performance for the first pass achieves less than 30% of recall for the HSBC logo and less than 10% for the SONY logo. In both cases, no false positives were detected.

The same figure shows the results for the two versions of the Extended Algorithm, which were introduced in section 2. There is a significant improvement in the recall performance in both cases. As expected, the version that propagates every frame achieves higher recall, detecting 18% of the HSBC logos and 25% of the SONY logos that were missed in the first pass. When only the first detected frame is propagated, the recall rate drops to 15% and 21%, respectively. However, as expected, there are no false positives for any of the logos when this second variant is employed. When every single frame is propagated, 10 false positives were detected for the SONY logo.

Along with the bar graphs for the two versions of the Extended Algorithm, two vertical bars are placed on the top of the columns that represent the aggregate results for the algorithm. These two bars indicate the percentage of misses from the first pass that cannot be achieved by the strategy employed in the Extended Algorithm. This situation occurs when a sequence of frames in which a logo is present does not have a match for any of its frames. In this case, the Extended Algorithm is not able to propagate frames and improve the performance. Using these two bars, it is easy to visually understand the maximum recall that can be achieved by the second pass. It is then clear that the second pass performs well and gets close to the maximum that it could possibly achieve.

It is clear that the results for the implemented system are not very good: the best-case recall is below 50%. However, this is mostly due to the poor performance of the first pass. The Extended Algorithm gives a performance enhancement in the order of 20%, capitalizing efficiently on the temporal redundancy inherent to a video source.

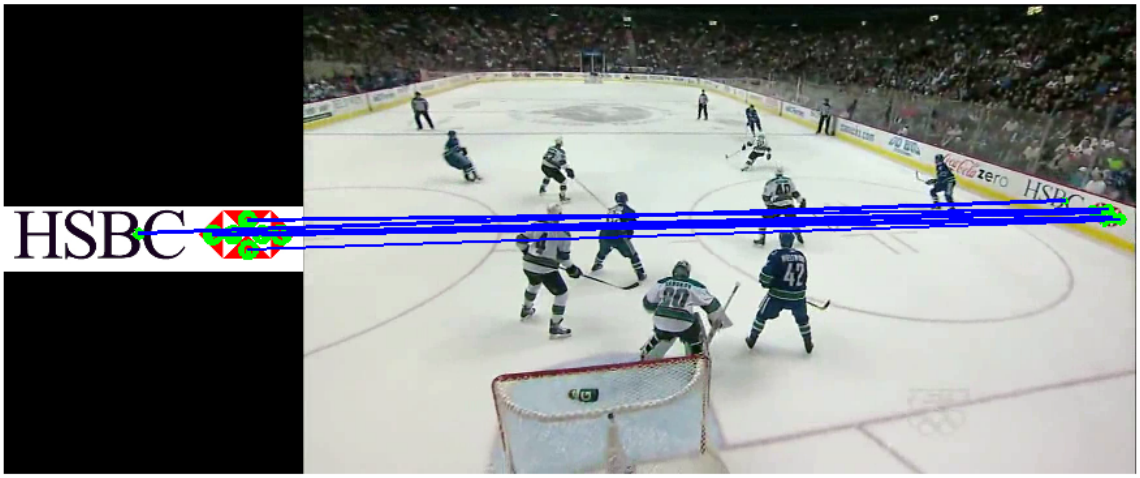

Fig. 4. Example of a correct match: frame 929. The blue lines connect matched features (green) from the images

Table 1. Recall results for the three algorithms. "Bas. Algorithm" refers to the Basic Algorithm. "Ext. Algorithm 1" refers to the Extended Algorithm with propagation for every frame. "Ext. Algorithm 2" refers to the Extended Algorithm with propagation only for the first frame.

Frame category	1		2		3		TOTAL	
HSBC	Count	Percentage	Count	Percentage	Count	Percentage	Count	Percentage
Ground truth	314	61.57%	146	28.63%	50	9.80%	510	100%
HSBC	Count	Recall	Count	Recall	Count	Recall	Count	Recall
Bas. Algorithm	112	35.67%	16	10.96%	0	0%	128	25.10%
Ext. Algorithm 1	175	55.73%	42	28.77%	1	2.00%	218	42.75%
Ext. Algorithm 2	164	52.23%	39	26.71%	0	0%	203	39.80%
SONY	Count	Percentage	Count	Percentage	Count	Percentage	Count	Percentage
Ground truth	71	46.41%	82	53.59%	0	0%	153	100%
SONY	Count	Recall	Count	Recall	Count	Recall	Count	Recall
Bas. Algorithm	0	0%	9	10.98%	0	-	9	5.88%
Ext. Algorithm 1	0	0%	46	56.10%	0	-	46	30.07%
Ext. Algorithm 2	0	0%	39	47.56%	0	-	39	25.49%

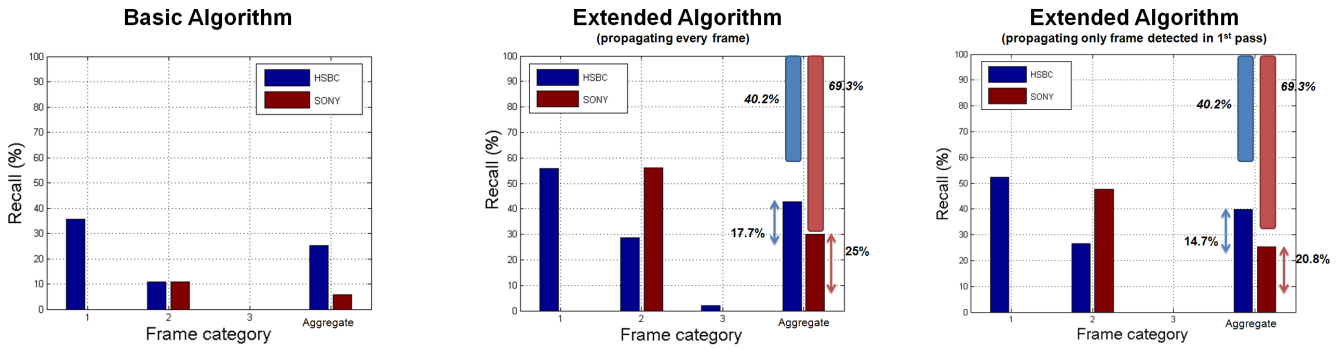

Fig. 5. Results for the Basic Algorithm and the two versions of the Extended Algorithm

4. CONCLUSION

In this paper we have presented a two-pass system to detect logos in a high-motion sports video. Two different approaches for using features from adjacent frames are explored, finding that there is a tradeoff between performance in recall and the appearance of false positives when propagating features from every frame or only from the frame that was originally detected. The bottleneck of the system is the first-pass. A higher recall from the Basic Algorithm would decrease the likelihood of an entire sequence of logo-containing frames to be missed. An improvement in this step would then result in an even greater improvement by the Extended Algorithm. Future work would mainly concern the first stage of the process, employing more sophisticated techniques, which could also consider temporal redundancy. An example could include a method to identify the quality of the individual frames and, from this information, adjust the thresholds to find more matching features with the database logo. Also, there is a lot of room for optimization of the algorithm, and this could be explored to further improve the system.

5. ACKNOWLEDGEMENTS

We would like to acknowledge Mina Makar, David Chen and Derek Pang for the frequent discussions that helped building a solid project.

6. REFERENCES

[1] D.S. Doermann, E. Rivlin, and I. Weiss, "Logo recognition using geometric invariants," *Proc. of the Second International Conference on Document Analysis and Recognition*, vol., no., pp. 894–897, Oct. 1993.

[2] A. Vedaldi and B. Fulkerson, VLFeat: An Open and Portable Library of Computer Vision Algorithms, 2008. `http://www.vlfeat.org/`

[3] D. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, vol. 60, no., pp. 91–110, 2004.

[4] M. A. Fischler and R.C. Bolles, "Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography," *Commun. ACM*, vol.24, no.6, pp. 381–395, June 1981.

[5] A. Bagdanov, L. Ballan, M. Bertini, and A. Del Bimbo, 'Trademark matching and retrieval in sports video databases," In *Proceedings of the international workshop on Workshop on multimedia information retrieval*, vol., no., pp. 79–86, 2007.

[6] M. George, N. Kehtarnavaz, M. Rahman and M. Carlsohn, "Real-time logo detection and tracking in video," *Proc. SPIE*, 77240B (2010); doi:10.1117/12.853598, 2010.

[7] S.Y. Arafat, S.A. Husain, I.A. Niaz, M. Saleem, "Logo detection and recognition in video stream," *Digital Information Management (ICDIM), 2010 Fifth International Conference on*, vol., no., pp. 163–168, July 2010.

[8] L. Ballan, M. Bertini, and A. Jain, "A system for automatic detection and recognition of advertising trademarks in sports videos," In *Proceeding of the 16th ACM international conference on Multimedia*, vol., pp. 991–992, 2008.

[9] R.J.M. den Hollander, and A. Hanjalic, "Logo recognition in video stills by string matching," *Image Processing, 2003. ICIP 2003. Proceedings. 2003 International Conference on*, vol. 2, pp. 14–17 Sept. 2003.

[10] A. Watve, and S. Sural, "Soccer video processing for the detection of advertisement billboards," *Pattern Recognition Letters*, vol. 29, no. 7, pp. 994–1006, May 2008.

[11] "Canucks Vs Sharks - Game Highlights - 03.18.10 - HD," http://www.youtube.com/watch?v=Ttt09PvHooU

Appendix

Breakdown of project preparation:

Andre: Implementations of the Basic and Extended Algorithm, literature review, poster and project report preparation.

Stephanie: Data collection and labeling, literature review, implementation of the Basic Algorithm, poster and project report preparation.