Int. Computer Symposium, Dec. 15-17, 2004, Taipei, Taiwan.

Robust Swimming Style Classification from Color Video

Wen-Hung Liao   Ming-Je Liu
Department of Computer Science
National Chengchi University
E-mail: whliao@nccu.edu.tw

Abstract *We present a robust method to classify swimming styles from live video based on the features extracted from the upper-body of the swimmer. In our approach, potential body parts are first extracted using a simple skin color model. The segmented regions are further analyzed to isolate the arm and shoulder blocks. Finally, a scoring system based on quantitative measures such as the slope, size, aspect ratio and the relative position of the body parts is constructed to carry out the classification. Several enhancements to the original scoring system are developed and tested. Experimental results demonstrate the validity and efficiency of our proposed approach.*

Keywords: Articulated Motion Analysis, Swimming Style Classification, Sports Video Analysis.

1. Introduction

The analysis of human motion is a research domain that has received considerable attention from the computer vision community in the past decade [1]. Automatic processing of sports video that involves single or multiple players is, in particular, a heavily investigated subject due to general public's interests in sport events. In general, there are two distinct purposes for analyzing sports video: 1) to organize and segment the raw footage into semantic units for easy indexing and retrieval, and 2) to perform qualitative or quantitative analysis of motion sequence to identify weak points and offer suggestions for improvement. The former has long been known to be a key step in building multimedia database. Video-content segmentation and highlight extraction has been an active research area [2]. More recently, leading international standard organizations (e.g., MPEG of ISO/IEC [3] and ATVEF [4]) have also started working actively on frameworks for organizing and storing such metadata. The latter is usually employed by a coach or professional athletes to improve skill and boost performance. For example, applied sports biomechanics concentrated on integrating the techniques from multiple disciplines such as physics, human anatomy, mathematics, computing and engineering to analyze movements of human motion in order to prevent injury and improve performance. A swimming sport bio-mechanist can enhance a competitive swimmer's performance by discovering the best breathing pattern, or examining whether the propulsion in swimming is due primarily to lift or drag. Appropriate suggestions can then be made accordingly. More generally, sports biomechanics involves the study of biomechanical properties of the movement of body parts, such as the arm moving cycle frequency, the angles of arm movement in each phase of a patterned motion.

A major characteristic of sports motion is that inherent structure exists as a direct influence of the game rule or pre-defined exercise sequence. We are not dealing with random movement. Instead, domain-specific knowledge can usually be applied to aid the analysis. For example, there are four swimming styles commonly swum in competitions. Three of them are regulated by *La Fédération Internationale de Natation* (*FINA, International Swimming Federation.*). They are: butterfly, backstroke, and breaststroke. A fourth competition is for unregulated styles and is called freestyle. During freestyle, it is possible to swim any style on this list. Due to the superior speed, most swimmers choose front crawl for freestyle competitions. For medley swimming, freestyle is any style except breaststroke, backstroke, and butterfly. As a result, the freestyle and front crawl are regarded as the same in our investigation.

In [5], we have reported a classification scheme to recognize swimming motion under the following constraints: 1) the swimmer is sited in a FINA Olympic standard pool with distinct lane ropes, 2) the video is taken with an above-water camera looking down as the swimmer is approaching, and 3) there is only one swimmer within the camera's field of view. While we have obtained satisfactory recognition results for all video clips that fit the above profile using the decision rules proposed in [5], we believe that the style classification system can be further enhanced. This paper attempts to extend and improve the performance of our formerly developed system. Specifically, we will formulate new classification rules and rebuild the associated scoring mechanism to achieve better recognition results. Extensive experiments will be conducted to validate our claim.

To recapitulate, our originally proposed approach consists of four stages, including color-based segmentation, connected component detection, regression analysis, and motion classification, as depicted in Fig. 1. The overall structure of the newly developed scheme remains the same. Nonetheless, we have made significant modifications to the 'motion classification' component to achieve superior outcome.

The rest of this paper is organized as follows. In section 2, we briefly review the framework of the classification system and summarize the key ideas employed in each stage. Section 3 describes the proposed enhancements to the original classification scheme. Section 4 provides the experimental results and performance comparisons. Finally, concluding remarks are given in section.5.

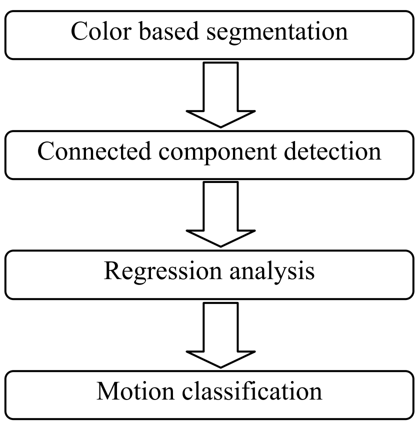


Figure 1. System framework.

2. System Framework

This section briefly reviews the four components of the classification system. A simple but effective segmentation method based on the hue value is first introduced. Candidate body parts are then isolated using connected component detection technique. Regression analysis is performed on the extracted regions. Finally, a scoring system based on the calculated data is developed to identify the type of motion.

Color video is employed to enable robust segmentation of the human body parts. The specific environment of the swimming pool has made this process quite simple as the background is largely blue, which happens to be on the opposite side of the orange-yellow interval in the hue ring. (Refer to Fig. 2). Fig.3 shows a typical frame of the input video clip and its hue component. Fig. 4 demonstrates the result of retaining the pixels that have a hue value between 0.3-1.5, i.e., potential skin pixels. We have found this simple technique to work quite well for most video clips. It is possible to refine the segmentation process by taking into account the fact that the lane ropes is usually bright while the projected shadow appears dark. In addition, morphological filtering is applied to clean up isolated or disconnected regions. In most cases, less than 10 blobs are left for processing in the next stage.

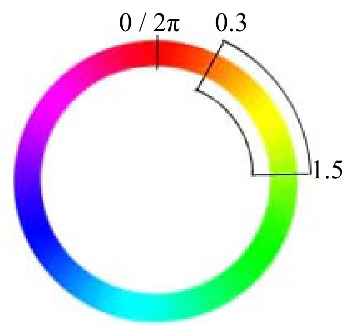
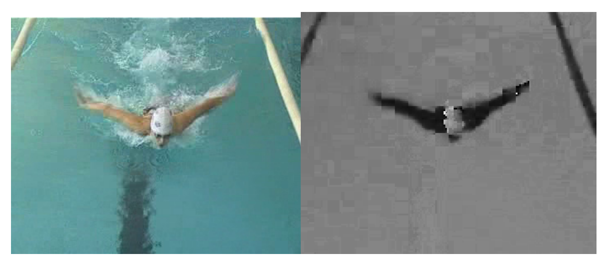
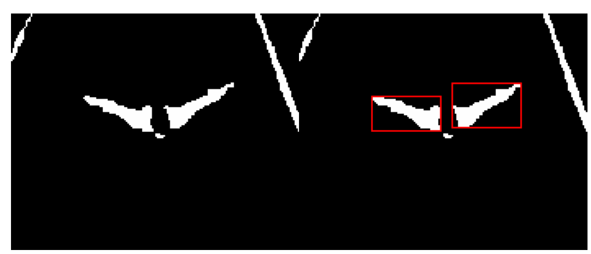


Figure 2. The interval corresponding to possible skin pixels.



(a) (b)

Figure 3. (a) Original color image (b) its hue component



(a) (b)

Figure 4. (a) Segmented regions (b) detected connected components.

After detecting the regions that could possibly be associated with the limbs, we need to identify the blobs that represent the arms. Depending on the swimming style being exercised, one or two dominant blobs may exist. We employ the algorithm documented in [6] to detect connected components. Components that cross the left/right boundary or have size smaller than a pre-defined threshold are excluded from further consideration. At most two blobs are retained for regression analysis.

The main purpose of performing regression analysis is to estimate the relative position of the limbs, which carries the key information in distinguishing different swimming styles [5]. We will rely heavily on the calculated slope of the blob to perform the classification. The correlation coefficient (R) measures how spread out the blob appears. It is used mainly to remove regions corresponding to the lane ropes, which usually possess large R values. Fig. 5 shows the detected blobs along with their regression lines.

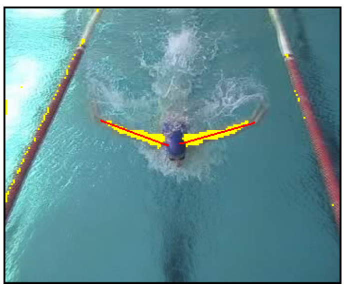


Figure 5. Connected components and their regression lines.

At this point, we are equipped with one or two dominant regions, the slope(s) of the regression lines, and the aspect ratio for each region. These features will constitute the basis for motion classification.

3. Motion Classification

As we have mentioned previously, swimming is a periodic motion that can be categorized into four stroke styles, namely, butterfly, backstroke, breaststroke, and freestyle. Each style depicts its unique combination of the movement of different body parts, including the arms, the legs, and the shoulder. In order to better characterize and analyze each motion style, it is desirable to gather as much data as possible from both the water surface and underwater. Our current settings, however, only supply the above-water video and thus only the upper body parts are visible in most image frames. Limited by such constraints, the arm movements will constitute the key clues in distinguishing the different swimming styles.

The resulting decision tree for classifying the motion sequence is depicted in Fig. 6. It differs from the original approach [5] in two important aspects: 1) a continuous scoring system is adopted to simplify the tree structure as well as increase the precision of the recognition results, and 2) a weighting factor is added to take into account the influence from the previous frames. Exploiting the relationship among successive image frames turns out to boost the performance of the classification system significantly.

The weight assignment algorithm works as follows: Originally, all four styles receive a weight value of 1. The weights are first updated at the *k+1* th frame according to:

$$\text{weight} = \frac{\text{total scores for the specific style in k frames}}{\text{accumulated scores in the previous k frames}} \quad (1)$$

The statistics obtained from our collection of swimming footage indicate that a complete cycle of the butterfly style takes about 28 frames. The estimates are 25 frames for backstroke, 30 frames for breast stoke and 20 frames for freestyle. Since it only makes sense to relate the motion patterns within one cycle, k should be less than 20. In practice, k is usualy set between 5 and 10 to account for dropped frames.

According to the decision tree depicted in Fig. 6, each frame in a video sequence may be associated with 0, 1, or 2[1] styles. Let *count[style]* denote the number of frames that have been assigned to a specific style in the past k frames. Based on the assumption that the swimmer will maintain a fixed swimming style, we may incorporate the rules shown in Fig. 7 to further adjust the weighting factor.

```
if (count[fly]>count[breast])
      weight[breast]*=(1-weight[breast]);
else if(count[back]> count[breast])
      weight[breast]*=(1-weight[back]);
else if(count[free]> count[breast])
      weight[breast]*=(1-weight[free]);
else if(count [breast]> count[fly])
      weight[fly]*=(1-weight[breast]);
else if(count [breast]> count [back])
      weight[back]*=(1-weight[breast]);
else if(count[breast]>count[free])
      weight[free]*=(1-weight[breast]);
```

Figure 7. Fine-tuning the weight.

[1] Butterfly and backstroke are indistinguishable in some frames.

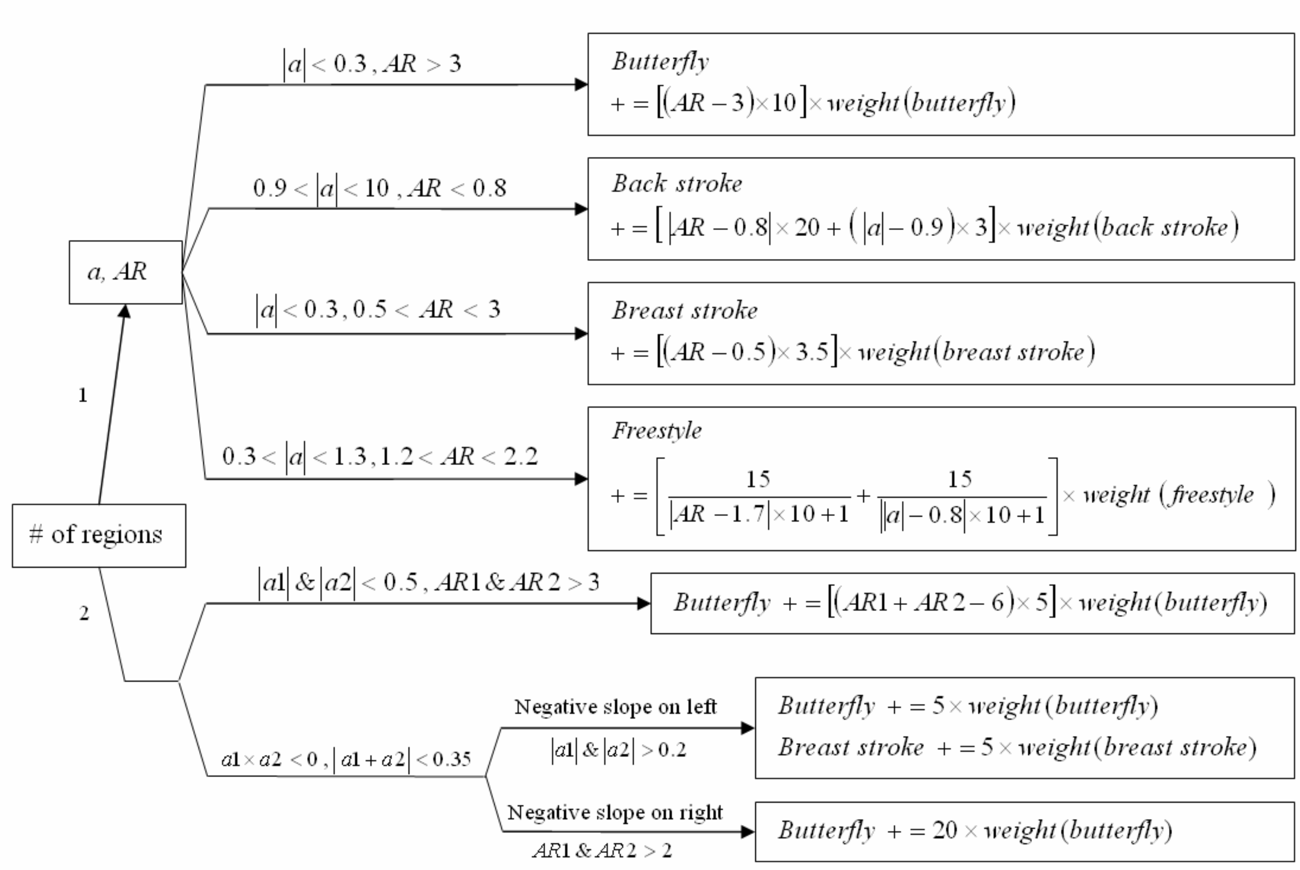


Figure 6. Linear decision tree and scoring system for swimming style classification.

It should be noted that we are not using a machine learning approach such as neural networks on the outset for specific reasons. Because swimming has pre-defined exercise sequences and the motion pattern is fixed for each style, the movements in each style usually will not differ significantly at every arm-movement cycle. Such domain-specific knowledge allows us to implement empirical rules to aid the analysis. Experimental results verify that the use of a rule-based system is sufficient. On the other hand, black-box approaches usually call for a large amount of training data, and is not guaranteed to offer better performance. Training time and collection of 'representative' samples raise other concerns.

4. Experimental Results and Discussions

In this section, we present experimental results applying the proposed method to classify motion patterns from video sequences in real-time. We also compare the performance of the different revisions to the original technique.

4.1. Experimental results

The database contains swimming footage download from http://swim.ee. About 50 video clips that fit the profile discussed in section 1 are selected. A total of 8552 image frames (approximately 285 minutes of video) are employed for the testing. We first apply the original technique to do the classification and compute the percentage of frames that have received correct labels. The results are listed in the second column of Table 1. We then initiate the first revision: using linear scoring system instead of the discrete one. The results are given in the third column. In most situations we have seen comparable or better recognition rates by incorporating continuous scoring. Column 4 summarizes the result of applying only the second revision: bringing in a weighting factor (with k=10). The improvements are much more pronounced in this case. The recognition rates have increased 8-20%. Combined effects of these two revisions are shown in column 5 of Table 1. Not surprisingly, the integrated approach has improved the overall recognition rates even further.

Table 1. Comparison of the recognition rates with different decision rules.

	original	(1) linear score	(2) weighting factor	(1)+(2)
Butterfly	64.04%	74.84%	78.18%	86.73%
Back stroke	88.42%	85.69%	96.85%	97.51%
Breast stroke	60.99%	57.64%	80.51%	78.77%
Freestyle	58.59%	63.87%	77.79%	93.18%
Average	68.01%	70.51%	83.33%	89.05%

The data in Table 1 are obtained by analyzing every frame in a video clip. To study how the proposed algorithm works on slower machines, we have also conducted experiments by 1) randomly dropping frames and 2) analyzing only 3 frames per second. The results depicted in Table 2 have clearly demonstrated the efficacy and robustness of the motion classification scheme.

Table 2. Comparison of the recognition rates with different frame rates.

	Analyze all frames	Randomly dropping frames	Analyze 3 frames/sec
Butterfly	86.73%	90.07%	82.02%
Back stroke	97.51%	95.98%	88.52%
Breast stroke	78.77%	77.62%	66.07%
Freestyle	93.18%	84.37%	68.36%
Average	89.05%	87.01%	76.24%

4.2. Discussions

With our proposed algorithm, we get 100% correct classification results when the swimming footage is analyzed frame-by-frame. (Using a majority vote principle, the classification will always be correct if the corresponding style receives over 50% of the total scores.) Repeated experiments indicate that if the frame rate is over 17 frames, all the classification results are still correct. However, when the frame rate has been reduced to 3, misclassification will occur. This is mainly due to the incorporation of the weighting factor, which exercises some form of 'inertia' in accordance with the properties of the previously analyzed frames. In the 3 frames/sec mode, the effect of noisy or misclassified frames may be amplified, resulting in adverse influence on the following video frames. Since butterfly and breaststroke are indistinguishable in some frames, the score may be distributed to both styles in a butterfly clip. In most breaststroke frames, we will not be able to see the swimmer's arms. Instead, the shoulder is detected, forming a rectangle which is very similar to the arm region in freestyle. As a result, the scores for breaststroke and freestyle are complementary. These phenomena are clearly depicted in Fig. 8.

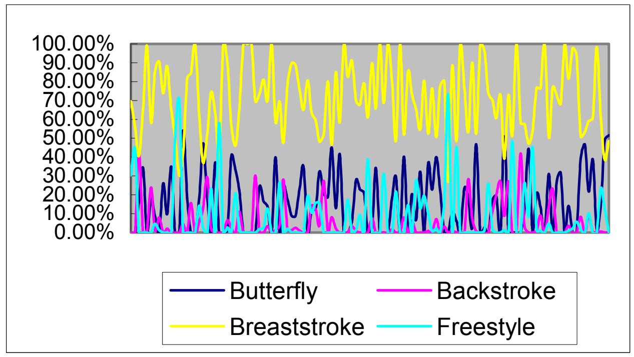


Figure 8. Statistics for breaststroke when analyzing 3 frames per second.

5. Conclusions

In this paper, we have presented an efficient and robust approach to classify swimming style using features extracted from the upper body parts. Based on the slope and aspect ratio of each candidate blob, the characteristics of various motion styles can be revealed. The proposed decision tree and scoring system prove to work well for video clips satisfying proper constraints.

Nonetheless, swimming videos come in many different flavors. The research reported in this paper resolves only a subset of all possible input patterns. For example, the current system will experience difficulties with video clips that are either too short or contain medley (mixed styles). Future work includes the detection of 'turn' to enable analysis of mixed swimming modes as well as the development of new rules to classify video taken from various view angles.

References

[1] J. K. Aggarwal, Q. Cai, W. Liao and B. Sabata, "Nonrigid motion analysis - articulated and elastic motion", Computer Vision and Image Understanding, vol. 70, no. 2, 1998, pp.142-156.

[2] H. Zhang, Y. Gong, S. Smoliar and S. Tan. 1994. "Automatic Parsing of News Video", Proc. of IEEE Conference on Multimedia Computing and Systems 45-54.

[3] MPEG-7: Context and Objectives (V.7). 1998. *I*SO/IEC JTC1/SC29/WG11 N2207, MPEG98.

[4] ATVEF http://www.atvef.com

[5] W. Liao, Zhung-Xun Liao, and Ming-Je Liu, "Swimming Style Classification from Video Sequences", 16th IPPR Conference on Computer Vision, Graphics and Image Processing, Kinmen, R. O. C., Aug. 17-19, 2003.

[6] R. C. Gonzalez and R. E. Woods, Digital Image Processing, Addison-Wesley Publishing, 2002.