Improved Maritime Target Tracker using Colour Fusion

Paul Westall, Peter O'Shea, Jason J. Ford
School of Engineering Systems
Queensland University of Technology Brisbane, Australia
{p.westall, j2.ford, pj.oshea}@qut.edu.au

Stefan Hrabar
Commonwealth Scientific and Industrial Research Organisation
Brisbane, Australia
stefan.hrabar@csiro.au

ABSTRACT

Searching for humans lost in vast stretches of ocean has always been a difficult task. This paper investigates a machine vision system that addresses this problem by exploiting the useful properties of alternate colour spaces. In particular, the paper investigates the fusion of colour information from the HSV, RGB, YCbCr and YIQ colour spaces within the emission matrix of a Hidden Markov Model tracker to enhance video based maritime target detection. The system has shown promising results. The paper also identifies challenges still needing to be met.

KEYWORDS: Colour Fusion, Hidden Markov Models, Mathematical Morphology, Maritime Search and Rescue.

1. INTRODUCTION

Human maritime search and rescue missions have always been a challenging task, and an element of chance is involved in the detection of survivors [1]. Humans become fatigued and complacent after long hours of searching, reducing the chance of finding survivors. We propose the use of machine vision to automate the location of human survivors lost at sea, and present an evaluation of colour space fusion within a Hidden Markov Model tracker framework to improve the performance of such a system.

Many countries have vast search and rescue regions. Australia's covers approximately 53 million square kilometres, which is equivalent to nearly one tenth of the Earth's surface. This is an enormous area to search and if robots, such as Unmanned Aerial Vehicles (UAVs) were employed, the probability of locating survivors would increase. Given this search capability, robots would be able to assist and aid present manned search endeavours by utilising UAVs as a task force-multiplier. This would allow the current search efforts to be more flexible and to respond with appropriate force in a timelier manner.

A small number of studies into the automation of maritime searches have previously been conducted, but all of these studies have restricted the search to small vessels (such as life rafts) and high visibility targets. Most notably, Sumimoto, et al [2, 3] have investigated the search for small bright orange life rafts in the ocean. Also, Toet [4] investigated the maritime search problem by choosing to fuse the morphological top-hat information from two different IR spectral frequency bands to reduce the effects of ambient noise in the image, while searching for approaching kayaks. Both of these investigations used images taken from static platforms looking out across the ocean, as from the bridge of a ship, searching for targets that are highly visible and largely above the surface of the water.

The investigation in this paper builds on previous work by the authors in [5]. That work used a morphological filtering operation followed by a Hidden Markov Model (HMM) based tracker to detect human heads in the ocean. The advancement presented in this paper is the incorporation of a data fusion technique to enhance the detection process. This fusion is accomplished by sourcing multiple data sets from different video colour spaces, similar to a technique used in biometric authentication [6], and combining them within the emissions matrix of the HMM.

This paper is organised in the following way: Section 2 contains a brief description of the problem and the constraints involved. Section 3 presents an overview of the proposed system, and Sections 4 and 5 detail the target detection and tracking phases of the system, respectively. The experiments performed are described in Section 6 and the results reported in Section 7.

2. PROBLEM DEFINITION

This paper specifically addresses the detection of a single person in the water during a daytime aerial search, in a scenario where the survivor is without a high contrast floatation or location device. The images are taken from a downward looking camera and the ocean surface currents are no faster than 2m/s [7].

To maximise the effective search corridor of a single camera the proposed algorithm searches for persistent point-like targets within the surveillance video. This feature of the search approach also has a bearing on the flight altitude that the system can be flown at, given the resolution and field of view (FOV) of the camera.

Searching for a person in the water is difficult because the target may only occupy 1-3 pixels in a dynamic and constantly changing environment, and the person is therefore not easily distinguished from the background. Furthermore, the target has limited time within the FOV of the camera – depending on the speed of the aircraft, the target maybe in view for as little as 3-5 seconds. Hence, a decision on the target's status must be made quickly to allow the operator to respond to the alarm.

3. SYSTEM OVERVIEW

In point detection applications it is common to use a pre-processing operation (such as spatial masking [8]) to reduce the ambient noise. Subsequently, various temporal tracking techniques can then be used to discriminate the true target from noise by exploiting properties of the target such as size, shape, colour and temporal dynamics.

Therefore, the proposed detection system comprises two main components: a point target detection phase (front-end) to reduce noise and identify candidate pixels, followed by a temporal tracking phase that uses *a priori* knowledge and historical data to discern true target behaviour among the various candidates (see Figure 1). This tracking incorporates the fusion of the data from the various different colour spaces.

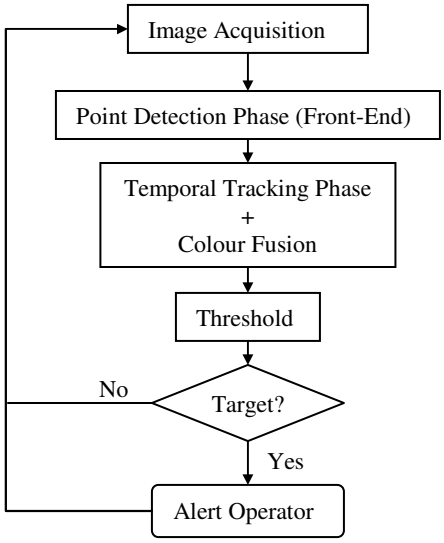

Figure 1. System Block Diagram

4. POINT TARGET DETECTION PHASE

Small targets are difficult to detect because they are not easily distinguishable from noise and/or clutter. In the proposed system, the automated search would be conducted at an altitude that would result in an average human head occupying only 1-3 pixels of the search camera's FOV, eliminating any shape information [9]. For these reasons we have employed a point detection technique from the area of Mathematical Morphology.

Morphology is based on using a structuring element B to perform two fundamental operations, dilation and erosion, on the target image A. Combinations of these operations create the open and close functions, described as follows:

$$\textit{Dilation} \quad A \oplus B = \left\{ z \middle| \left[(B)_z \cap A \right] \subseteq A \right\} \tag{1}$$

$$\textit{Erosion} \quad A \Theta B = \left\{ z \middle| (B)_z \subseteq A \right\} \tag{2}$$

$$\textit{Opening} \quad A \circ B = (A \Theta B) \oplus B \tag{3}$$

$$\textit{Closing} \quad A \bullet B = (A \oplus B) \Theta B \tag{4}$$

Morphological filtering of images (either Electro-Optic (EO) or Infra-Red (IR)) has proven effective for identifying candidate targets in the areas of aircraft collision avoidance [10] and multi-spectral IR target detection [4].

The particular morphological filtering implementation of the close-minus-open (CMO) technique presented in this paper is a consolidation of the filtering method employed by Casasent [11] and the filter application technique of Deshpande [12]. Four 1D slit-shaped structuring elements (one vertical, one horizontal, one on the leading diagonal

and one on the trailing diagonal) are applied at both the close and open steps. This quad-filter approach allows only targets which are compact in all directions (i.e. point-like) to be enhanced. Non-compact clutter, such as white caps and large floating debris, tend to be attenuated. It has been found that the basic CMO operation gave rise to images that do not correctly represent the zero mean nature of the image noise. This misrepresentation in turn degrades the effectiveness of the tracking phase of the target identification. The following "preserved sign" CMO method was introduced to overcome this problem [10]:

$$CMO_{signed} = (F_{in} - (F_{in} \bullet SE)) + (F_{in} - (F_{in} \circ SE))$$
$$CMO_{signed} = 2F_{in} - ((F_{in} \bullet SE) + (F_{in} \circ SE)) \quad (5)$$

where CMO_{signed} = Signed Output Image
F_{in} = Input Image
SE = Structuring Element

Using the above modification to the usual CMO technique reduces false alarms by an average of 20% [10].

5. TEMPORAL TRACKING PHASE

To enhance the identification of potential targets one can make use of all measurement information over a substantial time period. This approach is typically referred to as track-before-detect (TBD), and is a commonly used concept in radar technology. The approach allows for the identification of targets in high noise and high clutter. A threshold is imposed as a final stage to the system to grant target status to those candidates that have been successfully tracked over a significant number of frames.

For the data considered in this study the search for targets between consecutive frames is restricted to within a 5x5 pixel kernel (because of the target's dynamics within the maritime application). These limitations are described more fully in the next paragraph.

Assume the use of a 1024x768 pixel camera operating at 15 frames per second (fps), at a nominal search height of 150m above sea-level, and travelling at 150km/h [1]. The highest documented ocean surface current speeds (barring extreme environmental anomalies) are in the vicinity of 2m/s [7] which is approximately 7km/h. In addition to this surface current, the maximum swimming speed of a human is roughly 8km/h [13]. Aggregating these two velocities still does not exceed the maximum velocity allowed by our target dynamics model.

To see this, suppose the FOV of the camera in the direction of the aircraft x-axis is 60°. This makes each pixel equivalent to 0.226m (average human head diameter [9]). Thus, for a target to move at least two pixels, relative to the ground, between consecutive frames it would have to be travelling at a minimum speed of approximately 25km/h – a speed that is highly improbable for the type of targets of interest. As a result, for the problem of airborne maritime searches, the slow-moving nature of the target allows the discrete velocity space to be limited to ±1 pixel in both the x- and y- directions. An additional transitional buffer of one pixel is given in all directions to allow for any error in image correction due to the moving platform. This justifies the use of a 5x5 kernel to examine changes from frame to frame.

5.1. Hidden Markov Modelling

Hidden Markov Modelling is a powerful statistical tool involving stochastic processes that can be represented as underlying discrete-value Markov chain state processes that are partially observed through a sequence of measurements [14]. HMMs have found use in many areas, such as signal processing, speech recognition and document character recognition applications.

A HMM is characterised through the following:

- The set of Markov chain states $S = \{s_1, s_1, \ldots, s_N\}$, where N is the total number of valid states for the model; and
- The HMM parameter set $\lambda = (\pi, A, B)$, where:

π is the initial state distribution vector (also known as *prior probabilities*), e.g. π is the probability of state i at the arbitrary time t=0.

A is the state transition matrix (also known as the *transition probabilities matrix*), where $A = [a_{ij}]$, with a_{ij} being the probability of transition to state j given current state i.

B is the output distribution matrix (also known as the *emission probabilities matrix*), where $B = [b_{ik}]$, with b_{ik} being the probability or likelihood of observing feature k given current state i.

In our approach the number of states, N, used in the HMM is equal to the number of pixels of the input image (representing each of the possible locations of the target).

After being initialised as per (6), the HMM filter essentially acts as a recursive algorithm which evaluates the probability of a target being at each location, given all previous observations. The output of this recursion algorithm is thresholded to determine if a target is present. The algorithm to yield the decision statistic, δ, is presented below [14].

$$\textit{Initialise} \qquad \delta_1(i) = \pi_i \cdot b_{i,x_1}, \qquad i = 1,\ldots,N \qquad (6)$$

where N = Number of states

π_i = Prior probability of being in state s_i at time $t = 1$

$$\text{Re } \textit{cursion} \quad \delta_t(j) = \sum_{i=1}^{N} \left(\delta_{t-1}(i) \cdot a_{ij}\right) \cdot b_{j,x_t}, \qquad (7)$$

where $2 \le t \le T$ & $1 \le j \le N$

t = *current time step*

x_t = *observation at time* t

The likelihood of a transition from one state, or pixel, to any of the surrounding states within the 5x5 kernel is defined by a discretised Gaussian surface with its peak located on the current state.

5.2. Colour Fusion

As the most likely part of the survivor's body to be seen above the water is their head, a provisional database of hair colours was compiled that were then modelled and combined to form target probability density functions (PDFs) to be used in the HMM. The target PDFs were generated by sampling a variety of naturally coloured hair specimens (i.e. black, honey blonde, brown, red and white hair) while the background PDF was assembled by sampling image frames of the test data that did not contain valid targets and approximating this data using a Gaussian curve.

For the detection process the raw RGB image is transformed into the required colour space, either Hue–Saturation–Value (HSV), Luma–Blue-Chrominance–Red-Chrominance (YCC) and Luma–In-Phase–Quadrature (YIQ) images, and then the desired layer is extracted. The extracted layer is then mapped to its corresponding target/background PDF curve, which is fabricated from the provisional database.

Also incorporated into the emissions matrix is an augmented version of the information produced by the image processing front-end, defined in (8). The greyscale readings given by the front-end are mapped to an empirically defined normal cumulative distribution curve to enhance the point-like interpretation of the front-end filter as valid dim point-like targets can be somewhat disadvantaged in the typical filtering process.

$$\Phi_{\mu,\sigma}(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} \exp\left(-\frac{(u-\mu)^2}{2\sigma^2}\right) du \qquad (8)$$

where x denotes the pixel greyscale level.

In order to populate the emission probabilities matrix, B, the colour space likelihoods (derived from their corresponding PDF curves) are fused along with the remapped version of the front-end output image within the multi-dimensional feature space of the HMM, as shown in (9).

$$P(v_k|s_j) = \prod_{i=1}^{M} P(f_i|s_j), \quad 1 \le j \le N \qquad (9)$$

Where v_k denotes the observed feature space $v_k = (f_1, \cdots, f_M)$, and s_j represents an element in the state space $S = (s_1, \cdots, s_N)$, with M and N being the total number of elements within the respective spaces.

By assuming independence between feature spaces fusion of the difference information sources can be achieved in the manner stated above. It is expected that the resultant image will give a better representation of the observed environment, therefore reducing clutter and other sources of false alarms, and enhancing the ability to effectively identify plausible point-like targets.

6. EXPERIMENTS

Below we describe trials that were conducted on both simulation and real flight data to gauge the performance of various system configurations, and to determine the most effective approach. All sequences consist of 30 frames with a single target appearing throughout.

6.1. Simulated Data Compilation

A simulated ocean scene was created using Blender, an open-source 3D animation application. The scene was animated to generate image sequences similar to those that would be produced during an aerial search operation, and these were used to test the detection algorithms.

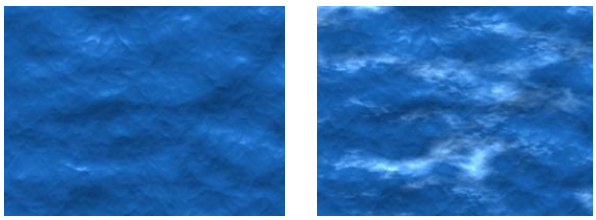

Figure 2. Blender Image Samples

Blender allowed us to vary wave height and speed, the colour of the ocean, the amount of white water as well as the reflections created by the sun. Two scenarios were

used – calm water and wavy white water (left and right images in Figure 2, respectively).

The scene was made to appropriate scale and the images shown in Figure 2 were created from a simulated camera height of 1000ft. A single human figure was also inserted into the scene with hair colour matching the modelled data used in populating the emissions matrix in the HMM.

Ten data sequences were generated from the two different sea states and five hair colours.

6.2. Real Data Acquired via a Flight Test

A flight test was conducted in a Cessna 172 over the beaches of the Gold Coast, Australia during June 2007. Images were captured from an altitude of approximately 500ft at 80 knots by a downwards-pointing camera mounted to the wing strut. A Point Grey Flea® camera fitted with a 185º FOV Fujinon YV2.2X1.4A-2 fisheye lens produced 1024x768 images at 15Hz. Camera pose (based on GPS and IMU) were also logged for each frame.

Although data was captured with a fish-eye lens, the region in the centre of the image has high spatial resolution without much distortion compared to the areas towards the periphery of the image. The centre portion of the original 1024x768 image was cropped to produce a 267x200 image. No further image rectification was performed.

A sample image frame from the collected data is shown in Figure 3. A single target is visible near the top-left corner (a surfer in a red rash shirt/wetsuit). This target remains within the cropped 267x200 image sequence for approximately 2 seconds, providing a suitable data set.

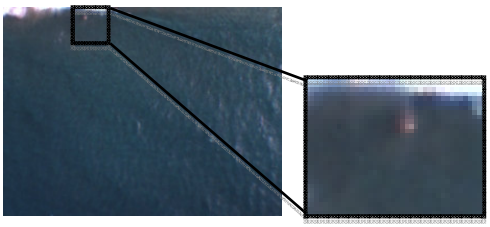

Figure 3. Flight Data Image Sample With Target

Note that for both the simulated and real data sequences, the images were post-processed to compensate for camera motion before being used by the detection algorithm.

6.3. Performance Metrics

A set of metrics were used to compare the performance levels of the system. The performance of the system was quantified using the following metrics:

FAR – False Alarm Rate is the average number of false alarms per image frame.

MDR – Missed Detection Rate is the average percentage (expressed as a fraction) of missed detections of true targets per image sequence.

FFTT – First Frame of True Target Detection.

FA length – False Alarm track length is the average number of consecutive frames that false alarms are considered targets.

Signal-to-Noise ratio of the image is calculated according to the following:

$$SNR = 10 \times \log_{10}\left(\frac{mean\left(\text{signal pixel intensity}^{2}\right)}{mean\left(\text{noise pixel intensity}^{2}\right)}\right) dB \quad (10)$$

7. RESULTS

Building on previous work we have selected the top performing colour space layers as identified within [5], namely: HSV1, HSV2, RGB1, YCC3 and YIQ2 (the trailing number denotes the specific layer of the triplet). All these layers will be used in the fusion segment.

Tests were performed with 5 different system configurations, each different system using one of the identified layers for use with the front-end, and all systems using the same fusion architecture. Representative subsets of the results are shown below.

7.1. Simulated Data Results

Key performance feature values of the tracking phase are defined in Table 1, below. These values have been averaged across all system configurations and input simulation data sets.

Table 1. Tracking Phase Performance

	FA Length	FFTT
No Fusion	0.6793	2.9578
Fusion	0.2077	2.1401

As shown, systems using fusion are capable of identifying the true target approximately 0.8 frames (0.05 seconds) earlier than systems that do no utilise fusion. Furthermore, as false alarms appear the systems with fusion are able to discern them three times quicker that those without.

Table 2. System Performance

	No Fusion		Fusion	
	MDR	FAR	MDR	FAR
HSV1	0.3623	0.2235	0.2481	0.1601
HSV2	0.3525	0.0882	0.0569	0.0000
RGB1	0.4864	0.2734	0.0822	0.0298
YCC3	0.0820	0.0000	0.0694	0.0000
YIQ2	0.0872	0.0000	0.0692	0.0000

A summary of the detection performance is given in the above table. The mean MDR and FAR, across all simulated sequences according to system type, are presented to illustrate the relative performance increase that fusion systems offer.

As shown in Table 2, the inclusion of fusion improves the overall performance of the detection system by reducing false alarms, while at the same time reducing missed detections. The colour space most greatly affected by the fusion method is HSV2. This layer experienced an 84% reduction in missed detections, as well as the total elimination of false alarms across all simulated sequences.

7.2. Real Data Results

Using the same system configurations, the detection algorithm was applied to real data sequences, producing the following plots.

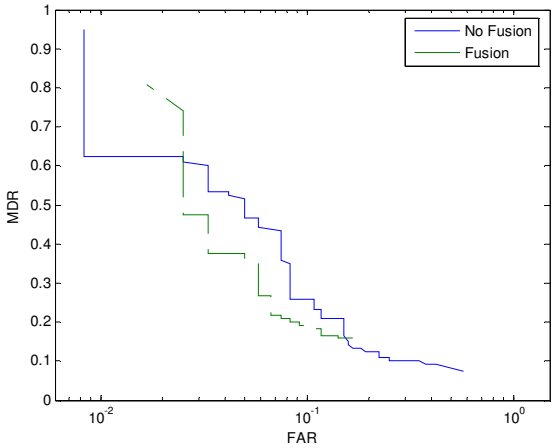

Figure 4. HSV2 Comparison

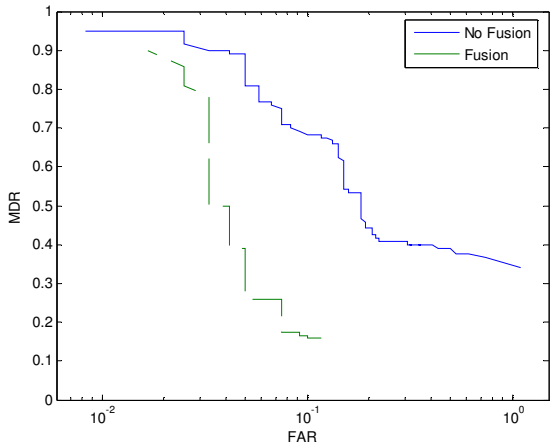

Figure 5. RGB1 Comparison

The effect of the incorporation of the fusion element into the system can be seen from the direct comparisons made in Figures 4 and 5. The cases shown here are depictive of the results obtained from the remaining colour space layers – albeit with varied results.

In the simulation data tests all the fusion systems achieved superior results, with many achieving zero false alarms. However, this consistent performance increase did not translate through to the results produced when applied to real data. We note that fusion adds significant improvement to the FAR and MDR of HSV2 and RGB1, marginal improvement to HSV1, but has a negative effect on the YCC3 and YIQ2 colour space performances.

Note however, that the background PDF used for the real data was different to that of the simulated data, as the colour distribution of the ocean between the sequences are not identical.

Results have given an encouraging indication to colour fusion's potential in improving visual tracking system performance. However, outcomes derived from the real data suggest that the output given by the system front-end has a strong bearing on the overall system performance. This will be the basis of further research.

8. CONCLUSIONS

These tests have shown that the fusion of colour information within the emission matrix of a HMM tracker, has the potential to significantly increase overall detection performance of the proposed system. While the outcomes of the simulated data have given strong evidence to the inherent advantage of colour fusion, the real data sequences have produced varying results.

Further investigation into the quality of the front-end output may be required in order to draw definitive conclusions as to the genuine effectiveness of colour fusion in this current application environment.

ACKNOWLEDGEMENTS

Flight data in this study was obtained courtesy of the Australian Research Centre of Aerospace Automation (ARCAA). Computational resources and services used in this work were provided by the High Performance Computing and Research Support Unit, Queensland University of Technology, Brisbane, Australia.

REFERENCES

[1] Australian National Search and Rescue Council, "National Search and Rescue Manual," Version 0501 ed. Australia: Australian Maritime Safety Authority, 2005.

[2] T. Sumimoto, K. Kuramoto, S. Okada, H. Miyauchi, M. Imade, H. Yamamoto, and T. Kunishi, "Machine Vision for Detection of the Rescue Target in the Marine Casualty," presented at Industrial Electronics, Control and Instrumentation, 1994. IECON '94., 20th International Conference on, 1994.

[3] T. Sumimoto, K. Kuramoto, S. Okada, H. Miyauchi, M. Imade, H. Yamamoto, and Y. Arvelyna, "Image Processing Technique for Detection of a Particular Object from Motion Images," presented at Industrial Electronics, 2001. Proceedings. ISIE 2001. IEEE International Symposium on, 2001.

[4] A. Toet, "Detection of Dim Point Targets in Cluttered Maritime Backgrounds through Multisensor Image Fusion," presented at Targets and Backgrounds VIII: Characterization and Representation, Orlando, FL, 2002.

[5] P. Westall, J. J. Ford, P. O'Shea, and S. Hrabar, "Evaluation of Maritime Vision Techniques for Aerial Search of Humans in Maritime Environments," presented at Computing: Techniques and Applications, 2008. DICTA '08.Digital Image, 2008.

[6] L. Nanni and A. Lumini, "Fusion of Color Spaces for Ear Authentication," Pattern Recognition, vol. In Press, Corrected Proof, 2008.

[7] K. Ouchi, "The Effect of SAR Bandwidth Ratio and Current Variation on Ocean Current Measurements by Along-Track SAR Interferometer," presented at International Geoscience and Remote Sensing Symposium, 1994. IGARSS '94. Surface and Atmospheric Remote Sensing: Technologies, Data Analysis and Interpretation, 1994.

[8] R. C. Gonzalez and R. E. Woods, DIGITAL IMAGE PROCESSING, Second Edition Ed. New Jersey: Prentice Hall, 2002.

[9] T. Partala, V. Surakka, and J. Lahti, "Affective Effects of Agent Proximity in Conversational Systems," presented at Third Nordic Conference on Human Computer Interaction, Tampere, Finland, 2004.

[10] R. Carnie, R. Walker, and P. Corke, "Image Processing Algorithms for UAV "sense and avoid"," presented at Robotics and Automation, 2006. ICRA 2006. Proceedings 2006 IEEE International Conference on, 2006.

[11] D. Casasent and A. Ye, "Detection Filters and Algorithm Fusion for ATR," Image Processing, IEEE Transactions on, vol. 6, pp. 114-125, 1997.

[12] S. D. Deshpande, M. H. Er, V. Ronda, and P. Chan, "Max-Mean and Max-Median Filters for Detection of Small Targets," presented at Signal and Data Processing of Small Targets 1999, Denver, Colorado, 1999.

[13] Infostrada Sports, "Swimming - World Record Progression Men 50m Freestyle," International Olympic Committee 8 July 2004.

[14] R. J. Elliott, L. Aggoun, and J. B. Moore, HIDDEN MARKOV MODELS: ESTIMATION AND CONTROL. New York: Springer-Verlag, 1995.