Australian Marine Mammal Centre

Final Report

(subclause 9 and Schedule Item 5 of the Funding Agreement)

Season

- **Title -** New Computerised Fluke Matching System for Humpback Whales
- **Chief Investigators** – Dr. Hendrik (Eric) Kniest, Daniel Burns and Associate Professor Peter Harrison
- **Organisation –** Southern Cross University & University of Newcastle

Activity Period – October 07 to September 08.

Table of contents

1. Activity Summary

A clear summary of approximately 500 words outlining the work undertaken and any significant findings (for publication on the Department's web site)
The new computer-aided fluke matching system 'Fluke Matcher' uses a unique multifaceted computer-based recognition system to overcome the overwhelming problems of manually matching photographs of humpback whale flukes in large catalogues. It provides a rapid and substantially improved method of analysing photo-identification data. The system is based on the unique characteristics of southern hemisphere humpback whale flukes, but could be adapted to suit northern hemisphere whales, as well as other marine mammal species. It significantly increases the efficiency of identifying individuals and finding resights in photo-identification catalogues of humpback whales. The system uses a wide range of criteria, based on multiple key features of humpback whale flukes that are normally utilized during manual matching methods, and additional computerised image-matching techniques, that produce a reliable matching system. Initial user input is needed to identify control points used for the transformation of the fluke image onto a common reference frame. The system then measures key features of the fluke, including parameters to describe the shape of the fluke, black and white pigment distribution in different regions of the fluke, and other distinctive features that enable identification. The system was developed and tested using a database of 117 photographs of humpback flukes with a total of 94 possible matches. Because of the broad scope of matching parameters used in the system, a number of different matching protocols can be used to rank potential matches. The optimum matching techniques resulted in 100% of the matches listed in the top 15 positions (out of 117), and 96% of the matches ranked in the top five. In this way the operator needs only to scan through a small percentage of the ranked images to find any possible match. The first step in the process of using Fluke Matcher to analyse photo-identification data is to measure or extract the data on each photograph. As part of this process it is important to transform the imaged fluke onto a common reference system. This is accomplished by the operator marking the position of five major control points on the photograph. These points include the most easily identified points, such as the fluke tips and central V-notch and two points on the leading edge. The next stage of the program ('Process Image') recalculates separate transformations for four separate districts across the fluke, but the transformed data are all defined in the one unique reference system. Angles and distances

are computed that define the shape of the fluke tips and the centre V-notch area, the thickness of the black band across the trailing edge, and the general shape of the fluke. The fluke is broken up into 18 regions and the percentage of black pigmentation is calculated for each of these regions. Next the operator can measure up to 5 different types of key features that help uniquely identify the fluke. These are classified as spot, line, area, damage and image features and are one of the most important aspects in finding any match for a fluke.

A user friendly graphical interface helps the operator progress through each phase. During the search process 'Fluke Matcher' matches each feature measured against each image in the database, giving each a score (0 to 100) for the probability that the feature properties are identical. An overall weighted Match Index (MI: 0 to 100) is then calculated for each image. The images are then ranked in order from the most likely match down to the least likely match and displayed in that order. The operator can then scan through the list to visually compare images and identify matches.

2. The Outcomes/Objectives

The degree to which the Activity has achieved the objectives

This project has resulted in a new computer-based photo-recognition matching system that efficiently identifies individuals and finds resights in photo-identification catalogues of humpback whales. 'Fluke Matcher' has reached all its objectives. It provides:
1) a user friendly interface to help guide the operator through each process of the measurement phase
2) a means to identify and measure unique key features that form the foundation of the matching process
3) a flexible matching technique to allow for strengths and weaknesses in the characteristics of the fluke and quality of the photographs
4) a graphical display to allow the user to scan through the ranked images and choose any match.

Most importantly, the overall findings from the tests show very promising results. A database of 117 fluke photographs was used to test and modify the search protocols. Only good quality, high resolution, photographs were used, to avoid any complications that may be associated with the use of poorer quality photographs. Using the best search methods for each fluke, all of the matches were listed in the top 13% of the list.

A large data set of 434 flukes representing 194 individuals was later entered into the system, taking an average of about 5 minutes for each fluke image to be entered. All photographs had passed the SPLASH protocols to ensure that poor quality photographs did not influence the results. About 90% of the matches ranked in the top 90/434 (20%) listed matches using the standard match search; 96% of matches ranked in the top 130 (30%). Although these photographs had passed the SPLASH protocols the quality of the photographs varied from low resolution scanned images to good quality images.

There still remain a small number of 'false-negative' matches (matches ranked in the lower end of the list) which are generally caused by photographs that have the two leading edge control points underwater; reducing the accuracy of the transformation process. Many of these photographs were low resolution, poor quality, and only showed the trailing edge of the fluke; but they still contained useful information for identifying the fluke. The matching results will be substantially improved if these photos are removed from the database, by setting higher standards for acceptable images; or developing additional matching techniques that overcome the problems associated with photos that have much of the fluke underwater.

3. Appropriateness

The appropriateness of the approaches used in the development and implementation of the Activity

The approach used to develop 'Fluke Matcher' has proven to be highly appropriate, resulting in a substantial increase in the capability and efficiency of analysing photo-identification data and therefore greatly improving data from photo-identification studies. The system does not rely on pure image matching techniques, but instead aims to substantially improve efficiency in manual matching methods. The system will greatly reduce the time needed to identify individuals and find resights in photo-

identification catalogues of humpback whales. Each fluke only needs to be entered into the database once and the structure of the database allows for rapid searching and matching. The database is also a convenient system for cataloguing fluke images and the system could be modified to list and display images based on certain criteria (fluke characteristics, time, date, location etc.).

4. Effectiveness

The degree to which the Activity has effectively met its stated objectives

Tests carried out using 'Fluke Matcher' show the system is effective in meeting its overall objective. The time taken to enter a new fluke into the database is around 5 minutes (depending on the quality of the photograph and the number of features that must be entered). Once this is done, the system improves on visual manual matching in a number of ways:

1) Each fluke needs only be entered into the system once (the image does not have to be rescanned, categorised or stratified each time a match is undertaken).

2) The search process effectively increases the potential for finding a match by reducing the total amount of images that needed to be compared and reduces some of the bias of manual matching.

3) All information (including fluke data, photo data and match data) is stored in a database structure that can be easily accessed and cross-referenced by researchers.

4) Any new image entered into the system can be quickly matched against any of the existing databases in the system. The time taken to search for a match in a database of about 500 images is only a matter of seconds. It may then take the operator a few minutes to scan through the ordered list to find any potential matches. This represents a very highly significant increase in the efficiency of matching humpback whale flukes, especially in very large catalogues.