Evaluating facial recognition for photographic-mark-recapture of four species of northern ungulates

by

Isobel Frances Gawn Ness

A thesis submitted in partial fulfillment of the requirements for the degree of

Master of Science

in

Wildlife Ecology and Management

Department of Renewable Resources
University of Alberta

Abstract

Estimating animal abundance is a key component of wildlife management and mark-recapture surveys are one of the most commonly used methods of obtaining population estimates. Photographic identification has recently been explored as a method of 'marking' individuals for mark-recapture surveys. It is most often used on species with spots, stripes of other unique markings however, in a few cases, species that do not have such obvious marks have been successfully identified using morphological measurements. There is a need for the development of non-invasive, affordable and accurate methods of censusing wide-ranging and elusive northern populations of ungulates. In this study, I tested a likelihood-based photographic identification method on 4 species of ungulates: muskox (*Ovibos moschatos*), Dall sheep (*Ovis dalli*), mountain goats (*Oreamnos americanos*) and mule deer (*Odocoileus hemionus*). False-rejection (FRR) and false-acceptance (FAR) error rates were identified for each species. These measures varied widely among species (i.e., FRR = 0–13%, FAR = 2–22%). Matching success was also determined for each species and ranged from 48% to 96%. Muskox (FRR=0% and FAR =3%) and sheep (FRR=4% and FAR=2%) had the lowest misidentification rates and the highest matching success rates (96% and 88%, respectively). Moderate results were obtained for goats (FRR=11% and FAR=6%) and deer (FRR=11% and FAR=0%) with matching success rates of 81% and 80%, respectively. An automated matching success rate was calculated based on the top-ranked photograph for each potential match and was compared to the observer matching success rate. The observer matching success rate was significantly higher for all analyses (t_9=7.2, $p<0.05$), indicating that the final subjective user choice step of the method was important. An observer bias test was conducted for deer and sheep and significant observer bias was found for deer, and affected matching success rates for both species pointing to the importance of observer training and/or experience in the species of interest and the use of the program. This study provides a proof-of-concept for the use of photogrammetric identification on sheep and muskox and lays the groundwork for future capture-mark-recapture studies on wild

populations by establishing misidentification rates and determining the effect of observer bias. Future work should address the stability of horn measurements over time and the feasibility of capturing useable photographs from remote cameras.

Preface

This thesis is an original work by Isobel Frances Gawn Ness. No part of this thesis has been previously published.

Acknowledgements

Many thanks go out first to my supervisors Fiona Schmiegelow and Tom Jung for their support and guidance. I would also like to thank the Yukon Wildlife Preserve in Whitehorse, Yukon, and Randy Hallock in particular, for allowing me to conduct research there and also providing me with access to the enclosures. Thank you also to the Large Animal Research Station in Fairbanks, Alaska for allowing me access and giving me a private tour. My family also made this project possible: Ray supported me, Charlie made me laugh, Number 2 kept me focused, and my parents babysat. Thanks.

Contents

Table of Figures

1.0 Introduction

1.1 Estimating Animal Abundance

Since the 1950s, attempts have been made to solve the problem of estimating animal abundance (Otis, 1978) and a wide range of methods exist to address the issue. Estimates of animal abundance inform most wildlife management decisions and allow managers to track the impacts of those decisions, as well as other influences such as climate change, development, and habitat loss. Generally, a population estimate is used to make inferences about total population, as counting every animal in a population is often not feasible. The simplest methods involve counting individual animals, either in random sample plots, or in line transects, and calculating population density estimates (Schwartz & Seber, 1999). Sampling is usually randomly distributed on the landscape to ensure representative samples are captured. However, in the case of species or populations that have aggregated distributions (like some birds or fish), adaptive sampling may be used. Adaptive sampling involves using knowledge obtained from initial sampling to inform future sampling locations (Schwartz & Seber, 1999). Estimating animal abundance using such methods can, however, be difficult for highly mobile, cryptic, or patchily distributed species.

Capturing, tagging, and marking animals has long been used by wildlife managers for population estimation through capture-mark recapture (CMR) surveys. In this case, a sample of animals are captured, then tagged or marked, and released. After allowing time for the marked and unmarked animals to mix, subsequent rounds of capturing and marking occur in which some of the marked animals are re-captured and some unmarked animals are also captured. This allows for an estimation of the population size based on the proportion of marked individuals that are recaptured (i.e., the ratio of marked individuals to total population size at the original time of sampling is the same ratio of marked individuals to total sample size in the second sample). This is called the Lincoln-Petersen index and is used for closed populations (Seber, 1982). If further sampling events occur then capture histories can be collected for each individual and more accurate population estimates can be derived.

Closed-population CMR relies on the following assumptions: 1) that the population is closed (i.e., that there are no births and deaths, and no immigration or emigration); 2) that individuals do not lose their marks over the course of the study; 3) that all marks are correctly read and recorded; and 4) that the marking or trapping does not affect the probability of the animal being recaptured (Otis et al., 1978). Implicit in assumptions 2, 3 and 4, is the importance of the relative permanency of the marks, their identifiability and their relative lack of influence on the animal in question. Marks that rub off, are lost,

or are unreadable may bias population estimates, as can marks that affect the survival or fitness of the animal or alter its future behaviour. Historically, capturing and marking animals have been invasive practices requiring handling individuals, and often permanently scarring or affixing tags or collars, which may be detrimental to their fitness or survival.

1.1.2 Invasive Sampling Methods

Traditional approaches of tracking individuals use invasive and costly methods such as marking and radio-collaring animals (Powell & Proulx 2003; Krausman et al., 2004; Cattet et al., 2008). Marking or radio-collaring large mammals most often requires that they be captured and anesthetized, which can cause injury or mortality. Documented rates of mortality vary by species; for example, based on long-term research in Scandinavia: moose (*Alces alces)* 0.7%, grizzly bears (*Ursus arctos)* 0.9%, wolverines (*Gulo gulo)* 2.8%, Eurasian lynx (*Lynx lynx)* 3.9%, and gray wolves (*Canis lupus*) 3.4% (Arnemo et al., 2006). Mortality can be caused directly from the capture process or anesthetic, or indirectly due to secondary effects such as stress from pursuit, separation from offspring, or trauma from traps (Arnemo et al., 2006). Mortality rates, however, are often under-estimated, due to scavenging, malfunctioning transmitters, emigration, or simply due to too short a post-capture monitoring period to confirm survival (Cattet et al., 2008). Moreover, mortality rates provide only part of the picture, as other negative sub-lethal impacts may occur, including injury (Jacques et al., 2009; Dechen-Quinn et al., 2014) or behavioural changes (Rachlow et al., 2014; Jung et al., 2019) caused by capture and handling, as well as physical and psychological stress caused by trapping, tagging and marking (Wilson & McMahon, 2006; Delehanty & Boonstra, 2009).

In a study of long term capture effects on black bears (*Ursus americanus*) and grizzly bears, Cattett et al. (2008) found that capture and handling affected movement rates and body condition with effects directly proportional to the number of times an animal had been captured. Captured and marked white-tailed deer (*Odocoileus virginianus*) fawns had lower survival rates than unmarked fawns, possibly due to the marking reducing their cryptic camouflage, as well as abandonment by does (White et al., 1972). Offspring abandonment is a common problem resulting from capture and handling, and has been documented in ungulates in particular. Capture-induced abandonment of moose (*Alces americanus*) calves was 18% in one study (DelGiudice et al., 2015), and 28% in another; and ranged from 22% for caribou (*Rangifer tarandus*), 19% for pronghorn (*Antilocapra americana*), 15% for mule deer (*Odocoileus hemionus*), 11% for elk (*Cervus canadensis*), and 4% for white-tailed deer, among other ungulate species (Livezey, 1990).

Chemical immobilization can also affect the behaviour of some animals. In male bighorn sheep (*Ovis canadensis*), an increase in the frequency of dominance fights was observed post-capture, and captured males lost these fights, which affected their subsequent social status (Pelletier & Festa-Bianchet, 2013). Chemical immobilization and radio-collaring of mountain goats (*Oreamnos americanus*) was found to have a negative effect on their reproduction and survival (Cote et al., 1998). Moose were found to have lower reproductive rates following chemical immobilization from helicopters (Ballard and Tobey, 1981), while immobilization of pregnant moose in late winter can cause an increase in post-natal mortality (Larsen & Gauthier, 1989). Chemical immobilization and capture has also been shown to cause short term changes in movements and behaviour (Northrup et al., 2014; Rode et al., 2014; Becciolini et al., 2019; Jung et al., 2019).

The method of marking animals may also have detrimental effects on individuals. Marking methods include brands, implanted tags, external tags, or scarring (Walker et al., 2011). Hot iron branding is the most effective way to mark many marine mammals, such as elephant seals (*Mirounga leonina*), as tag loss can be very high; however, it can cause large weeping sores and infected wounds, and result in the animals being in poor physical condition in some cases (Caiafa et al., 2005). Flipper banding of seabirds has been used since the 1950s to mark individuals, but has also been linked to increased energy expenditure and foraging time, lower survival rates and severe flipper damage (Sherley et al., 2010). Moose calves marked with ear transmitters were found to have a significantly higher mortality rate than unmarked calves and calves marked with ear tags (Swenson et al., 1999).

Radio-collaring is one of the most common methods of monitoring wildlife and can be used to collect information about home range size, movement patterns, and behaviour. However, the collars can have negative impacts on animals and neck lesions have been documented across a range of species, including mule deer, bighorn sheep, and howler monkeys (*Alouatta palliata*) (Krausman et al., 2004; Hopkins & Milton, 2016). Global Positioning System (GPS) and Very High Frequency (VHF) collars placed on mule deer for one year were found to cause skin loss and ulcer formation, and similar injuries were found when collars were placed on bighorn sheep (Krausman et al., 2004). Skin loss and more serious subcutaneous tissue and muscle damage injuries were found when ball chain radio collars were deployed on female mantled howler monkeys (Hopkins and Milton, 2016). While advances in collar technology and deployment methods reduce these kinds of injuries, they still occur in some cases and avoiding collaring altogether is the most effective way of avoiding the associated detrimental effects. Hence, there is a need for less invasive sampling techniques to mark individuals for CMR studies and to

address issues of animal welfare, as well as those of measurement bias introduced by the methods themselves. GPS collars, for example, are prone to frequent failures and are costly, which can lead to reduced sample sizes and poorer statistical inference (Hebblewhite & Haydon, 2010). While invasive methods of sampling, tracking and marking are predominant, new non-invasive approaches are being developed to track animals and identify individuals for population counts and demographic studies (Waits & Paetkau 2005; Bolger et al., 2012).

1.1.3 Non-Invasive Sampling Methods

In northern Canada, the development and application of non-invasive sampling methods for wildlife is also necessary to address cultural sensitivities to invasive sampling methods. The Inuvialuit and Dene believe that live wild animals should not be handled, and that it is disrespectful to affix any man-made object to their bodies (Byers, 1999). The Heiltsuk First Nation of BC similarly voiced concern over the radio-collaring of wolves in their traditional territory (Darimont et al., 2008). These concerns and beliefs are shared across many northern indigenous communities (Byers, 1999), and are not necessarily unique to First Nations. The evolution of co-management practices and policies, such as those that have been created as a result of Yukon First Nation land claims agreements, and the need for consultation with First Nations without such agreements, means that addressing these concerns is vital.

Non-invasive methods of identifying individuals, such as genetic sampling (Waits & Paetkau 2005), and photogrammetry (Berger, 2012), are being developed to monitor and study populations while reducing disturbance to animals and costs to researchers and managers. Non-invasive genetic sampling methods can reveal information about populations and individuals from the extraction of genetic information from hair, feces, or other DNA sources. Genetic sampling can, however, be limited in its application due to the rapid degradation of DNA, high genotyping error rates, contamination, relatively low success rates (Waits & Paetkau, 2005), and high cost. Genetic sampling methods can thus benefit from cross correlation or validation with other methods (Anile et al., 2014; Drechsler et al., 2015).

1.2 Photographic Mark Recapture

Photography has been used to document wildlife occurrence and behaviour almost since its advent. Remote deployment of cameras has also recently precipitated a revolution in the collection of data in the field of wildlife biology (Cutler & Swann, 1999; Dertien et al., 2014). Remote cameras have been primarily used for the collection of data on behaviour, predation, activity patterns, habitat use and occupancy (Burton et al., 2015). The use of remote cameras for the collection of demographic information has been slower, but recent advances in image processing software have allowed the

development of programs to uniquely identify individuals of a population. Photographic mark-recapture (PMR) has recently gained attention as a method of estimating animal abundance (Lubow & Ransom, 2009; Bolger et al., 2012; Tancredi et al., 2013; Alonso et al., 2015; Schmidt et al., 2017). Photographic mark-recapture is similar to other forms of CMR modelling; however, it uses photographic identification to "mark" individuals and record capture histories, as opposed to traditional mark-recapture monitoring, which uses physical marks and requires physically recapturing individuals (Lubow & Ransom, 2009). Photographic mark-recapture demands that individuals be amenable to being photographed (either remotely, while captured, or while free-ranging), have variable traits that can be measured, and that these traits remain stable over the period of study (Bolger et al., 2012). Applying PMR to a new population requires identification of the phenotypic traits to be measured, development of a method of capturing the traits in photographs, and determination of the length of time over which the traits are stable. The application of PMR to northern species would increase the tools available to researchers and wildlife managers, and could allow for the population estimation of wide-ranging species that are otherwise costly or difficult to monitor.

The application of PMR methods involves identification of individuals from photographs, which requires quantifying unique phenotypic characteristics (Kelly 2001; Goswami et al., 2007; Sacchi et al., 2010). The method was initially developed on species with spots, stripes, or other unique markings, such as tigers (*Panthera tigris*; Karanth & Nichols, 1998), zebras (*Equus quagga*; Foster et al., 2006), whale sharks (*Rhincodon typus;* Meekan et al., 2006), marbled salamanders (*Ambystoma opacum;* Gamble et al., 2008) and giraffes (*Giraffa camelopardalis tippelskirchi*; Bolger et al., 2012). Three-dimensional computer-aided matching has been used on Serengeti cheetahs (*Acinonyx jubatus*) with a high degree of accuracy (100% when poor quality and bad angle photos were removed; Kelly, 2001). Camera trap studies on tigers demonstrated that individuals could be accurately identified by their stripe patterns, and the resulting data could be used for CMR modelling (Karanth & Nichols, 1998). Zebras can similarly be differentiated based on stripe patterns (Foster et al., 2006). Marbled salamanders also have unique dorsal patterns that have been used in mark-recapture studies (Gamble et al., 2008). Colouration of the pelage on the chest of Asiatic black bears (*Ursus thibetanus)* and sun bears (*Helarctos malayanus*) has been captured via specially designed remote camera setups: the method was found to be highly reliable, cost effective, and easy to use, for determining population status (Higashide et al., 2012; Ngoprasert et al., 2012). Wolverine also have distinctive coat colouration on their chest which can be captured with specially designed remote camera setups and used to "mark" and identify individuals (Magoun et al., 2011). Similarly, African penguins (*Spheniscus demersus*) have unique chest plumage patterns, and a

fully automated program was developed for population monitoring of this species (Sherley et al., 2010). Whisker spot patterns also produce a unique 'fingerprint,' and have been used to identify individual polar bears (*Ursus maritimus*; Anderson et al., 2010) and Australian sea lions (*Neophoca cinerea*; Osterreider et al., 2015).

Photographic identification studies have been conducted on a large range of marine animals (Table 1). In many cases, they provide ideal candidates, as they are often the subject of long-term studies, which result in more individuals being 'captured' and thus increases the reliability of the method (Hillman et al., 2003). In addition, marine tourism and citizen science databases have been established to help document their locations and populations. A prime example of a charismatic, highly photographed marine species is the *Wunderpus photogenicus* octopus, which can be uniquely identified from white spots on its dorsal mantle and is a popular target of tourists' cameras (Huffard et al., 2008). In harbor seals (*Phota viculina)*, pelage markings, flipper scars and tags were used as unique identifiers (Hastings et al., 2008), based on a similar method developed for gray seals (*Halichoerus grypus*; Hiby & Lovell, 1990; Hiby et al., 2013). Mediterranean monk seals (*Monachus monachus*) have also been identified using unique pelage patterns (Forcada et al., 2000).

Another type of computer-assisted matching program relies on a string-like representation of the curvature of dorsal or pectoral fins (Araabi et al., 2000; Gope et al., 2005), and a similar multi-curve matching method was developed for the ears of African elephants (*Loxodonta spp*.; Ardovini et al., 2008).

1.2.1 Computer-Assisted Photographic Identification Programs

Many programs have been developed for computer-assisted photographic identification, some of which are species-specific, but many of which can be applied to multiple species and are generally free and open sourced. ExtractCompare (http://conservationresearch.org.uk/), for example, is an interactive photo matching software that uses 3D body models to automatically identify animals from their individual markings. It was originally developed for gray seals (Hiby & Lovell, 1990) but has been used on other species such as harbor seals and wildebeest (*Connochaetes taurinus*; Hastings et al., 2008; Morrison & Bolger, 2012). Misidentification rates for this method varied (Table 1) from 8% for wildebeest (Morrison & Bolger, 2012) to 14% for gray seals (Hiby & Lovell, 1990). Wild-ID (https://envs.dartmouth.edu/people/douglas-thomas-bolger) uses a Scale Invariant Feature Transform (SIFT) operator to reduce preprocessing of the images and allow the use of a greater range of images taken at varying angles. It has been used for the identification of wildebeest and Masai giraffes (*Giraffa*

camelopardalis tippelskirchi) but can theoretically be used for any distinctively patterned species (Bolger et al., 2012; Morrison & Bolger, 2012). Both programs require some preprocessing of the photographs and subsequent user selection from the top-ranked matches presented by the program, however ExtractCompare requires three preprocessing steps, while Wild-ID requires only one (Morrison & Bolger, 2012). Many marine mammals can be identified by their fin shapes, and a multi-species program called FinScan was developed to identify bottlenose dolphins (*Tursiops truncatus*), dusky dolphins (*Lagenorhynchus obscurus*), spinner dolphins (*Stenella longirostris*), long-finned pilot whales (*Globicephala melas*) and white sharks (*Carcharodon carcharias*), as well as sperm whale (*Physeter microcephalus*) fluke photographs (Hillman et al., 2003). The program performed well for dolphins and sharks with the correct match appearing in the top-ranked position 50% of the time and in the top 3-4 matches 75% of the time; however, species-specific adjustments are required for use of the program on the two whale species. These programs are computer-assisted and are thus particularly useful in dealing with large databases, which are becoming increasingly common. A program based on an astronomical pattern matching algorithm was developed to deal with a large database of photographs of whale sharks in Australia, based on the spatial arrangement of spots on their skin which achieved 90% matching success rates (Arzoumanian et al., 2005). To further automate and simplify the process another software program was applied to whale sharks called the Interactive Individual Identification System (I^3S) that can be used to identify individuals based on distinct spot patterns, achieving a 93% matching success rate (http://www.reijns.com/i3s/; Speed et al., 2007). HotSpotter is another software for the automated identification of patterned species that has been used on Grevy's zebras (*Equus grevyi*), plains zebras (*Equus quagga*), giraffes (*Giraffa spp*.), leopards (*Panthera spp*.) and lionfish (*Pterois spp*.; Crall et al., 2013). It achieved 100% matching success for giraffes, lionfish and jaguar, 95-98% for Grevy's zebras and 99% for plains zebras. AMPHIDENT was developed for the identification of great-crested newt (*Triturus cristatus*) from their ventral spot patterns and achieved a false rejection rate of 2%; it was cross-correlated with a genetic study and found to have better recapture rates (Drechsler et al., 2015). The authors posit that the program could have a much wider applicability to a broad range of species with highly variable and discriminative patterns such as sand lizards (*Lacerta agilis*), Galapagos marine iguanas (*Amblyrhynchus cristatus*), adders (*Vipera berus*), Near East fire salamanders (*Salamandra infraimmaculata*), and Pyrenean mountain brook newt (*Calotriton asper*).

Spots, stripes and other patterns are recognizable, variable and unique; however, other measurable features or groups of features can similarly be used to identify individuals of species that don't possess such conspicuous 'marks'. Pumas (*Puma concolor*), for example, do not possess spots or stripes, but

may be uniquely identified by scars, kinked tails, tail-tip coloration, and other subtle marks, with an average agreement between pairs of investigators of 79.3% (Kelly et al., 2008). Coyotes (*Canis latrans*) have similarly been identified using colouration, banding and other pelage variation characteristics with 100% agreement between five observers, albeit when comparing only 3 individuals (Larrucea et al., 2007). Physical characteristics combined with spatial locations and group sizes have been used to temporarily "mark" grizzly bears in a CMR survey that produced abundance, density and occupancy estimates (Schmidt et al., 2017). Identification of many species of deer using fur characteristics, in combination with other identifiers such as the presence or absence of antlers or fawns, was deemed successful in a camera trapping census in Mexico where 97% agreement between two independent observers was achieved (Soria-Diaz & Monroy-Vilchis, 2015). However, less subjective methods have also been developed such as photogrammetry and facial recognition.

1.2.2 Photogrammetry

Photogrammetry (taking measurements from photographs) has been successfully applied to a number of species globally for the estimation of population condition and individual fitness (Krause et al., 2017; Berger, 2012), age (Bergeron, 1992; Flinn, 2010; Gee et la, 2014), as well as the unique identification of individuals (Merkle & Fortin, 2013). While such measures are typically collected from dead or live-captured individuals, non-invasive methods have been developed. Parallel lasers, for instance, have been used to measure horn size in ibex (*Capra ibex*; Bergeron, 1992), and drones have been used to measure body mass and estimate body condition in pinnipeds (Krause et al., 2017). Manual photography and remote camera traps in particular have become increasingly useful in this regard. Photogrammetry has been used to measure body size variation in moose and its effect on juvenile survival, and head size variation in muskoxen (*Ovibos moschatos*) in relation to environmental variation (Berger, 2012). The measurement of antler size from photographs of white-tailed deer has been used to accurately estimate their age (Flinn, 2010). While useful, this type of photogrammetry is limited to the measurement of morphological traits which can help estimate a population's condition or even its demography, but not population size. The application of photogrammetry to the identification of unique individuals is an obvious next step in the evolution of the use of photogrammetry.

1.2.3 Facial Recognition

Facial recognition, as developed for humans, involves automated programs that take morphological measurements of human faces. Eigenfaces is a computer-assisted photographic identification program based on principal component analyses of human faces (Turk & Pentland, 1990). It was subsequently applied to elephant seals (*Mirounga leonina*) with 100% effectiveness (Caiafa et al., 2005). Other human

facial recognition software programs have been extended and applied to chimpanzees (*Pan troglodytes*) for use in mark-recapture studies, with false rejection rates ranging from 12.5% to 17.7%, and false acceptance rates ranging from 3.5% to 13.5% (Loos & Ernst, 2013). A more recent and accurate facial recognition of chimpanzees was developed using a deep learning artificial intelligence that achieved 92.5% recognition accuracy (Schofield et al., 2019). A similar feature extraction program was applied to lemurs with a 98.7% success rate (Crouse et al, 2017). Identification of loggerhead sea turtles (*Caretta caretta*) has also been achieved with a 99% success rate using facial photographs of their scale patterns and an identification tree (Schofield et al., 2008). Common bottlenose dolphins have also been the subject of a facial identification study. It was not computer-assisted but was conducted by 27 biologists looking at folders of left and right side photographs and was compared to identification done by dorsal fin. The results suggest that the dolphins are symmetrical, such that they can be identified by right or left-sided facial images, and that the facial traits are stable over long periods of time (up to 5 years; Genov et al., 2017). Facial identification in wildlife is otherwise relatively uncommon (Table 2). However, a method has been developed for identifying individual bison (*Bison bison*) from photographs of their faces, using measurements of their horns in combination with other morphological traits (Merkle & Fortin, 2013).

Morphological measurements taken from photographs of 33 known free-ranging bison from Prince Albert National Park, Canada, using the program ImageJ, were first converted into ratios to standardize the measurements across photographs taken from different distances (Merkle & Fortin, 2013). The ratios were then compared between pairs of individuals and a probability statistic, and likelihood scores were generated to determine the prospect of a match. The top five matches were presented to the user for final selection of the matching individual. The program relies on high quality photographs from which details such as hair colour differences and horn patterns can be distinguished, as well as head-on shots with little deviation from a forward-facing angle. While the use of photographic identification has been rapidly expanding, most of the work has relied on pattern recognition and not on photogrammetry of morphological traits. As such, this method is unique in that it uses morphological measurements of the horns and faces of the ungulates. Importantly, though, the method relies heavily on horn measurements. The method has not yet been used on ungulates without horns (i.e., cervids). The extension of this method to antlered species, and to other species of bovids, would expand the applicability of the method and provide a proof of concept for the use of facial recognition in ungulates.

The main objective of my study is to test and apply a photographic identification method to four species of ungulates with the goal of providing a non-invasive method that may be used to estimate animal abundance. My specific objectives are to 1) test Merkle and Fortin's (2013) likelihood-based photographic identification method that was developed for bison, to other species of bovids including sheep, goats and muskox; 2) modify the method such that it can be tested on a species of cervid (mule deer); 3) test a modified method to determine the relative importance of horn measurements for unique identification of a bovid; 4) test the importance of the final subjective step of the method to determine its relative importance compared to a completely automated version of the method; and 5) determine the effect of observer bias on the method. I predicted that this method would be applicable to other species of bovids due to the similarity of the morphological facial features available to be measured. The application to a species of cervid proved more challenging due to their lack of horns; however, deer possess other facial landmarks that may be measurable, and facial colour patterns that may allow for identification. The final step of the method requires that the user choose the right match from the top-ranked matches provided by the program; while this step likely improves the accuracy of the method; it also adds a significantly time-intensive and subjective step. Quantifying its value for different species, as well as the resultant misidentification rates, may be an important consideration for future application of the method. In addition, observer bias may be particularly problematic in photo-identification studies where there are varying degrees of experience and of subjectivity. The effect of observer bias is a factor in considering the use and application of the method and quantifying if it will be valuable for potential users, particularly in relation to different species.

Methods to uniquely identify individuals may be achieved a variety of ways, and has been developed for many species (Table 1 & 2). Most species that do not possess spots, stripes, or other distinct 'fingerprints', require more creative solutions to be uniquely identified. Northern ungulates fall into this category and are also species that would benefit from the development of a non-invasive sampling method like photographic identification. Large and remote areas with wide-ranging, elusive species are often hard to survey accurately and the high costs of traditional invasive methods can limit their use. Additionally, the development of non-invasive sampling methods would help to address the cultural sensitivities of northern First Nations to invasive sampling methods. Photographic identification based on morphological measurements of faces and horns is a method that has worked for bison and is likely applicable to other species of bovids: it may also be adaptable to cervid species, further expanding the use of the method.

2.0 Methods

To test Merkle and Fortin's (2013) likelihood-based matching algorithm on different species of ungulates, I first collected data at the Yukon Wildlife Preserve (YWP) in Whitehorse, Yukon between September 2016 and May 2017, and at the Large Animal Research Station (LARS) in Fairbanks, Alaska, in July 2017. The YWP is a not-for-profit organization that seeks to promote research into non-invasive wildlife conservation techniques. The YWP has captive populations of 12 species of mammals that they aim to maintain in as wild a state as possible. Some of the animals were born in captivity while others were brought to the preserve from the wild. Many of the muskox at LARS came from farms and the research aims of the facility are mainly aimed at the raising of muskox for commercial purposes: thus most of the animals have had their horns cut off making them un-useable in this study. The habitats at both facilities are fenced, and in some cases quite large, often making photography of the entire population challenging.

I took photographs of 19 male and female mountain goats, 26 female Dall sheep, 10 male and female muskox, and 30 female mule deer. The goats and muskox were not sexually segregated in photographic sessions as the populations were small (19 and 10, respectively) and they are less sexually dimorphic than the sheep and deer. The sheep and deer were segregated by sex in their enclosures and had large enough populations that they could be sampled separately. Females were chosen because, in the case of sheep, they had a larger population of females than males at the YWP. In the case of deer, females were chosen because they are antlerless and had a large population for sampling. The sessions were conducted so that the maximum number of animals could be photographed and taken note of separately, as many of the animals were not tagged or otherwise individually identifiable. The methodological process is depicted in a flow chart (Figure 7).

Multiple photographs were taken of each individual capturing head-on shots and various angles deviating from head-on shots (up to 20°). Acquiring multiple photographs of each individual was required to be able to calculate photo error and to determine the program limits for using acute angle photographs. To adequately capture all of the individuals, deer were photographed on 6 different occasions, goats on 7 different occasions, sheep on 9 different occasions, and muskox on 3 different occasions at the YWP (and 1 occasion at the LARS). The photographs were taken with a Nikon D5100 Digital SLR Camera (Nikon, Tokyo, Japan) with an AF-S Nikkor 18-55mm lens and 16.2 megapixels and with a Fujifilm Finepix S3400 Digital Camera (Fuji, Tokyo, Japan) with a 28X Zoom and 14 megapixels. Two cameras were used because some individuals required more zoom (achievable with the Fuji) and

some required more speed (achievable with the Nikon); having two cameras also helped to avoid the problem of running out of batteries during a photographic session. Photo quality was subjectively comparable between the cameras.

2.1 Morphological measurements

Distances between selected morphological (i.e., facial) landmarks in the photographs were manually measured (in pixels) using the program ImageJ (Abramoff et al., 2004). Initially 9–10 measurements similar to those used by Merkle and Fortin (2013) for bison were measured including: distance between the horn tips, length of each horn, width of the base of each horn and distance between the eyes. Subsequently, up to 20 other measurements were made for deer, sheep, and goats, these included: the distance between the inner and outer corner of the eyes, length and base width of the ears, length of the nose, width of the nose, and septum, and a variety of measurements between the mid and endpoints of these measurements. Twenty measurements were used to account for the potential for increased false rejection rates with a larger population and a higher degree of photo error (Merkle and Fortin, 2013). To control for varying distances and angles between the photographs, each measurement was then standardized by converting it into a derived ratio. In the case of muskox, the measurements used were analogous to those used by Merkle and Fortin (2013; Figure 1 and Table 3), as were the derived ratios. Ratios 1–8 were calculated by dividing measurements M1–M8 by M9 (the measurement with the least variability). Ratio 9 is based on the ratio between M1 and M2: the widest point of the horn width and the tips of the horn width (Table 4).

The measurements used for deer were different than those for bovids, because cervids lack horns and their antlers are shed and regrow annually, rendering them an undiagnostic trait over short time periods. Twenty measurements were taken of the face and ears of the deer (Figure 2 and Table 5). These measurements were chosen based on traits often used in human facial recognition analyses, the morphological features available for measure in a head-on photograph, and those used by Merkle and Fortin (2013). The ratios were calculated by dividing each measure by M1 (the measurement with the least variability) and ratio 1 is based on the ratio between M4 and M3, the widest part of the nose and the narrowest part of the septum (Table 6).

The initial sheep measurements taken were also based on those used by Merkle and Fortin (2013); however, 10 measurements were taken instead of 9 because a distance between horn base measurement was available that was not possible with bison or muskox, where their forelock obscures their horn bases in most cases (Figure 3 and Table 7). The derived ratios were calculated in the same

way as for muskox and bison: ratios 1–7 and 9–10 were calculated by dividing measurements M1-M7 and M9-M10 by M8 and ratio 8 is based on the ratio between M1 and M2 (M1-M2/M2) (Table 8). For a subsequent sheep dataset, the same 20 measurements that were used for deer were used, and the horn measurements that were used initially were not used (Figure 4 and Table 9). Ratios were calculated the same as the deer ratios (Table 10).

The goat measurements were also modelled on the measurements used for bison, muskox, and sheep; however, due to the shape of their horns (i.e., they are relatively straight compared to bison, muskox and sheep), a measurement at the widest point of the horns and the corresponding width of each horn was difficult to confidently obtain. To compensate for this, a measurement at the width of the nose at the nostrils and the length of nose were added (Figure 5 and Table 11). The ratios were then similarly calculated such that M1–M4 and M6–M10 were divided by M5 (the measurement with the least variability) and ratio 5 is based on the ratio of the outer horn base width to the inner horn base width (M8–M2/M2; Table 13). A subsequent method of analysis was attempted for goats based on that used for deer. Twenty measurements were taken of the faces of the goats and the horn length measurement and width of each horn base were included (Figure 5 and Table 12). The ratios were calculated by dividing each measure (M2–M20) by M5 (the measure with the least variability). As above, ratio 1 is based on the ratio between M8 and M2 (Table 14). A third goat analysis was tried in which 18 of the aforementioned 20 measurements were used, leaving out the horn length measurements, in an attempt to determine the relative importance of horns for goat identification. The dataset, measurements and ratios were otherwise the same (Figure 6, Table 15 & Table 16).

2.2 Photograph error and measurer error

Photograph error (caused by differences in the angles and distance of the photographs) was calculated by measuring 4–5 photographs each of 5 different animals and then calculating a mean standard deviation (SD) of each of the ratio. Measurer error (caused by human error or variation when taking the measurements in ImageJ) was calculated by the same observer measuring the same photograph 5 times and repeating this for 10 individuals of each species. The mean SD was then taken for each ratio and used to calculate the weights (see Tables 9–12). The weights were calculated as weighted means such that the ratio with the smallest mean SD was 1, with each other weight valued at less than 1, scaled based on their respective sizes. The SDs and weights were inputted into the *MatchImage* package in R.

For each analysis, the population SD of each ratio was calculated and compared to the photo error SD of each ratio. The photo error SD was divided by the population SD to determine the percent photo error

of each ratio. These were then averaged to calculate the average photo error for each analysis. The degree of photo error was categorized as low=10%, medium=30% and high=50%, expressed as a percent of the population SD, as done earlier by Merkle and Fortin (2013).

2.3 *MatchImage*

MatchImage is a package written by Merkle and Fortin (2013) for use in the Program R (available at https://r-forge.r-project.org/R/?group_id=1628). It uses a likelihood approach based on a Gaussian distribution and uses the morphological measurements to calculate a similarity score between an unknown photograph and a known photograph (at least one photograph in the database must be specified as known). To calculate the similarity score, for each pair of measurements (of the two photographs being compared), a density distribution is estimated based on the 2 ratios and their respective SDs; the density value is then multiplied by the weight. The sum of all the values (for each measurement) is then divided by the maximum possible similarity score (if 2 photographs had the exact same measurements). The similarity score is calculated as follows:

$$s.scr = \sum_{i=1}^{n} f(x_n|\mu_n, \sigma_n) \times \omega_n$$

where

$$f(x_n|\mu_n, \sigma_n) = \frac{1}{\sigma_n\sqrt{2\pi}} exp\left(-\frac{(x_n - \mu_n)^2}{2\sigma_n^2}\right)$$

where n=number of measurements used in the analysis, *f*=the probability density function that x_n is the same value as μ_n (where x_n=the measurement value of the known photograph and μ_n=the measurement value of the unknown photograph), given a Gaussian distribution with SD=σ_n, and ω_n=the weight of each measurement.

To conduct an analysis, the user compiles a database of measurement values, of which at least 1 must be specified as known. From this database, the program calculates the similarity scores and compares the unknown photograph to each known photograph in a pairwise manner. The user is presented with the unknown photograph and the closest potential matches and their similarity scores (up to 5), and must choose the correct match or specify otherwise.

Sets of photographs were analysed with the program using the 10 measurements developed for bison (Merkle & Fortin, 2013) for each bovid species (muskox, goats, and sheep). Datasets were chosen from

the database of photographs captured, based on quality and representation such that each individual animal was represented from 1-6 times in a set and the quality was such that identification could feasibly be achieved. Photographs in which the head was turned >20 degrees from head-on (determined by estimation from photographs) were excluded, as were those that were blurry or obscured. For deer, the above method could not be used due to a lack of horns, therefore novel measurements were used from the outset. The chosen set of 20 measurements were also used on subsequent analyses of goats and sheep: for goats as an attempt to lower misidentification rates and for sheep to determine whether horns are necessary for the method to work. Subsequent analyses were also conducted using different sets of photographs, in some cases by increasing the sample size for a more representative analysis, and in other cases by reducing it to determine a more accurate misidentification rate. To improve misidentification rates for deer, obscured, blurry, or badly-angled photos were removed from the databases and a re-analysis was conducted. In some cases, they were also segregated by season (i.e., sheep and deer).

2.4 Misidentification Rates

To evaluate the performance of each analysis, a false rejection rate (FRR) and false acceptance rate (FAR) were calculated. Two kinds of false rejections occurred: program false rejections (pFR) occurred when the program failed to produce a match in the top 5 potential matches, and human false rejections (hFR) occurred when a match was produced in the top 5 but was deemed by the user to be a non-match. These were combined into a total FR (tFR). False acceptance (FA) occurred when a match was falsely specified by the user to be a match. If a match was deemed to be a FA, a FR was not also considered for that possible match. False rejections result in an overestimation of population, while false acceptances result in an underestimation of population. The FRR was calculated by dividing the number of tFRs by the number of photographs in the database. The FAR was calculated by dividing the number of FAs by the number of photographs in the database. To determine whether a match was true or false, manual identifications were conducted on each photograph in each dataset and they were assigned a unique identification. The manual identifications were obtained by rigorously analysing each photograph manually, using data available from when the photographs were taken, and by analyzing additional photographs available from the database taken during the same session (i.e., side and full body photographs of the same individual). Photographs that could not be confidently identified in this manner were rejected from the dataset. This meta-data was not included in the datasets used for analysis to ensure it did not bias results.

2.5 Matching Success

The matching success obtained for each analysis was calculated by subtracting the total number of false rejections and false acceptances from the total number of potential matches (i.e., the number of unknown individuals in each dataset) and then dividing by the total number of potential matches. The total number of potential matches are a function of the number of photographs in the database minus the number of known individuals specified for each dataset.

2.6 Automated Matching Success

To determine the effect of the final subjective step of the method, an automated matching success was calculated for each analysis (i.e., what the matching success would have been had the program's top-ranked match been chosen automatically). The automated matching success was calculated by counting the number of true matches in the top-ranked spot and dividing it by the total number of potential matches. The matching success was then compared to the automated matching success rate and a t-test for paired two sample means was calculated to determine whether the difference between the 2 values was significant.

2.7 Observer Bias

To determine the effect of observer bias on the method (i.e., the last step in the process), an experiment was conducted to determine agreement among and between observers. Eight biologists identified deer and sheep from top-ranked images. They were given a brief explanation of the method and how to distinguish between individuals of each species and they were presented with the same datasets and photographs as those previously tested. The session was conducted with the top-ranked matches displayed on a screen at the front of the room. The observers were given brief instructions on how to identify each species. While the observers were all trained biologists, they had no prior experience matching photos of deer or sheep. The accuracy rate of each observer was calculated and an average accuracy rate for each species was calculated. The among-observer agreement was then calculated using an intra-class correlation coefficient (ICC). The ICC is calculated by dividing the variance among individuals (s_a^2) by the total variance (s^2; Hayes and Jenkins, 1997; Bell et al., 2009).

$$ICC = (s_a^2) / (s_a^2 + s^2).$$

In general, ICC values less than 0.5 indicate poor agreement and values between 0.5-0.75 indicate moderate agreement (Koo & Li, 2016). The ICC analysis was completed using package *irr* in R (R Core Team 2017).

3.0 Results

The results for all analyses conducted are summarized in Table 17. The misidentification rates ranged from a FRR and FAR of 0% and 2% for muskox and 4% and 2% for sheep, respectively, to a FRR and FAR of 28% and 9% for goats. Matching success rates similarly ranged from 96% for muskox to 48% for deer.

3.1 Muskox

The results of the muskox analysis were the most successful of all the analyses conducted. An analysis of muskox photographs obtained at both the YWP and the LARS was conducted on 31 high-quality photographs of 16 individuals. The 10 measurements used were analogous to those used for bison (Merkle and Fortin, 2013) and the ratios calculated are outlined in Table 4. Fifteen individuals were identified of an actual 16 and the FRR and FAR was 0% and 3% (due to one false identification), respectively. The matching success rate was 96%, while the automated matching success was 46%.

3.2 Sheep

For the sheep analyses, high success rates were achieved when the databases were segregated based on seasons, and an analysis conducted that did not use horn measurements had lower success rates. The first sheep analysis was conducted on a set of 76 photographs of 32 individuals collected on 3 days in September and October of 2016 and 1 day in May 2017. The 10 measurements used were analogous to those used by Merkle and Fortin (2013) and the calculated ratios, SDs and weights are listed in Table 8. The number of individuals was overestimated: identifying 41 individuals when there were 32. The FRR and FAR was 11% and 1%, respectively. The matching success rate was 80%, while the automated matching success rate was 52%.

A second sheep analysis was conducted using the same set of measurements on a subset of the photographs used for the first analysis, but included only the photographs from May 2017 and were thus segregated by season. It was conducted on 45 photographs of 19 individuals of which 21 were identified with a FRR and FAR of 4% and 2%, respectively. The matching success rate was 88%, while the automated matching success rate was 77%.

To determine the potential utility of the 20 measurements used on deer that do not rely on horn measurements, a third sheep analysis was conducted. These measurements did not include the horns as landmarks. From a set of 110 photographs of 26 individuals taken on seven separate days over a period between September 2016 and May 2017, the program identified 62 individuals and had a FRR of 25% and a FAR of 5%. The matching success rate was 62%, while the automated matching success rate was 28%.

3.3 Goats

Moderate results were achieved for goats when the highest quality photos were used and the highest number of measurements including horn measurements were used. The first goat analysis conducted used 67 photographs of 19 individuals from photographs taken on 4 separate days between September and November of 2016. The 10 measurements used were analogous to those used by Merkle and Fortin (2013) on bison and the calculated ratios, SDs and weights are listed in Table 13. Forty individuals were identified of an actual 19 and the FRR and FAR was 24% and 6%, respectively. The matching success rate was 59%, while the automated matching success rate was 37%.

A second analysis on goats using the 20 measurements that were used for deer, and also including 2 horn measurements, was conducted as an attempt to reduce the misidentification rates. The calculated ratios, SDs and weights are listed in Table 14. The second goat analysis used 47 photographs of 19 individual goats and the program identified 26 individuals: the FRR was 11% and the FAR was 6%. The matching success rate was 80%, while the automated matching success rate was 22%.

A third goat analysis used the same set of 47 photographs as for the second goat analysis, however only 18 measurements were included: the 2 horn length measurements were not used. As a result the program identified 31 individuals of an actual 19 and the FRR was 28% and the FAR was 9%. The matching success rate was 59%, while the automated matching success rate was 15%.

3.4 Deer

Results from the deer analyses were poor except in an analysis using a very small number of the highest quality photographs, in which case, moderate results were achieved. The first analysis of deer photographs was conducted using 60 photographs of 30 individuals taken on 4 separate days between October and November of 2016. Twenty measurements were taken of the faces of the deer: the calculated ratios, SDs and weights are listed in Table 5. Thirty-six individuals were identified of an actual 30: the FRR was 13% and the FAR was 22%. The matching success rate was 48%, while the automated matching success rate was 25%.

A second deer analysis was conducted on the same group of photographs in which poor quality photographs were removed including any partially occluded images, and any in which the deer had turned ears. The second deer analysis used 50 of the photographs from the first analysis and the same set of 20 measurements. Thirty-five individuals were identified by the program of an actual 30 individuals with a 22% FRR and a 12% FAR. The program matching success rate was 58%, while the automated matching success rate was 13%.

A third analysis of deer was conducted using a smaller dataset of 18 photographs of 8 individuals, representing the largest number of known individuals with high quality photographs that was captured in one photographic session (taken on October 19, 2016). The analysis identified 10 individuals of an actual 8 individuals and the FRR was 11% and the FAR was 0%. The program matching success rate was 80%, while the automated matching success rate was 20%.

3.5 Photo Error

Average photo error was calculated for each analysis conducted and is a measure of the calculated photo error SD as a percentage of the population SD. The calculated photo error SD was divided by the SD of each measurement for the total population. This was calculated for each measurement taken and then an average was calculated for each analysis. Average photo error ranged from 18% for sheep to 89% for goats (Figure 7). In general, as photo error increases FRR/FAR also increase.

3.6 Automated Matching Success

The automated matching success rates were calculated for each analysis conducted and were compared to the program matching success rates. The results from all analyses are plotted in Figure 8. A t-test for paired two-sample means was calculated to determine whether the difference between the 2 values was significant. The program-estimated accuracy rate was significantly higher than the automated matching success rate for all analyses (t_9=7.2, $p<0.05$).

4.7 Observer Bias

The observer bias session included 8 untrained observers who achieved matching success rates ranging from 23-42% with a mean of 32% ± 6% (SD) for deer. The matching success rate for sheep ranged from 73-96% with mean of 85% ± 7% (SD). The ICC or the inter-rater agreement was 0.142 (95% CI=0.033-0.339) for deer and 0.544 (95% CI=0.392-0.71) for sheep, indicating poor agreement between observers for deer and moderate agreement between observers for sheep (Koo & Li, 2016).

5.0 Discussion

5.1 Species differences in accuracy of facial recognition

The matching success obtained for muskox and sheep was 96% and 88%, respectively. In combination with the relatively low misidentification rates (FRR=0% and FAR=3% for muskox and FRR=4% and FAR=2% for sheep), a photogrammetric approach shows promise for use in PMR studies of these species. The matching success and FRR/FAR fall within the upper range of success rates obtained using other methods on other species (Table 1). For example, with a similar sample size of 42 individuals, humpback whales (*Megaptera novaeangliae*) were identified by their fluke patches with a FRR of 17%

(Ranguelova et al., 2004), and a population of 50 common wall lizards (*Podarcis muralis*) were identified by their ventral scales with a FRR of 2% (Sacchi et al., 2010). Many studies assume that FAR=0% (Bolger et al., 2012), and while this assumption held true for the analysis of muskox, it did not for sheep. This assumption may, however, be met if only the highest quality photographs are used, as the false acceptance in this case occurred on a slightly blurred photograph in which the distinctive horn rings were not as visible. The small population, high quality photographs and distinctive horns of muskox resulted in a very high accuracy rate of the method on muskox. In the case of sheep, the captive population also has very distinctive horns and were easily photo-captured (they spend time near the fence resting) so that I had a large database of high quality photographs to work with.

The moderate accuracy obtained for goats was somewhat surprising given the fact that they have horns and that the analysis was conducted on a small, captive population. The best result was obtained for the analysis using the 20 facial measurements, which also included horn length. The ten measurements used for goats in the first analysis differed from those used by Merkle and Fortin (2013) on bison because they did not have the requisite horn curvature to allow several horn measurements to be taken where horns were widest. Thus a different set of measurements was used for goats that relied less on horn measurements and instead used other morphological features such as the width and length of the nose (Figure 5). Despite having a small population and having horns, only moderate success was achieved, and the lowest FRR and FAR were 11% and 6%, respectively. The matching success rate was 80% which is in the low- to mid-range of published misidentification/matching success rates (Table 1). For example, for the matching of elephant ear-edge nick patterns, 83% true positives were achieved in highly cluttered and noisy images (Ardovini et al., 2008). Using images of grey seals in a software called ExtractCompare, an FRR of up to 33% was incorporated into a PMR study which produced relatively accurate population estimates (Hiby et al., 2013). The identification of individual chimpanzees was considered successful when a FRR of 12.5-17.7% and a FAR of 3.5-13.5% was achieved. It should be noted, however, that in all of these cases much larger sample sizes were used which would likely also affect the results achieved here. For a larger dataset, higher misidentification rates would be expected (Merkle &Fortin, 2013).

The database of goat photographs was also of a lower quality than the photographs of the other species due to the distance from which many were taken (resulting in blurriness) and badly-angled photographs, also caused by the difficulty of photographing the goats from behind a fence at a distance. Some examples of the poorer quality photographs are depicted in Figure 12. Approximately 15% were at a

bad angle (either looking downward or at > 20 degree angle from head-on) and 40% were blurred such that their horn rings were not visible. The low to moderate success rates achieved on goats may also be partly due to the high degree of morphological similarity between goats' horns, as well as the relative smoothness of their horns, as compared to sheep, muskox and bison. Mountain goats achieve >95% of their total horn growth by the time they are 4 years old (Cote et al., 1998), and thus variability in horn length beyond this age may be negligible. It is expected that with higher quality photographs, better results could be achieved. The low degree of variability between individuals' faces and horns meant that the program often failed to produce a correct match in the top 5 matches (4 out of 5 false rejections were program false rejections). In addition, they do not possess the differences in colouration of the face that deer and sheep have, as well as the highly visible horn patterns (often despite blurriness) that the sheep have, which often aid in successful matching during the final subjective step (there were 3 false acceptances and 1 false rejection due to human error).

Extending the method to a species of cervid (deer) without being able to rely on horn or antler measurements proved more challenging and resulted in high misidentification rates. The analysis of deer photographs had high misidentification rates for the first two analyses: FRR=13% and FAR=22% for the first analysis, and FRR =22% and FAR 12% for the second. The first analysis was conducted on 60 photographs of 30 individuals across seasons and the second analysis was conducted on the same set of photographs, but obscured and turned ear photographs were removed (Figure 10). This resulted in the FAR dropping from 22% to 12%, but had an opposite effect on the FRR which went from 13% to 22%. Additionally, the photo error rates were very high for both of these analyses: 77% and 74%. The matching success for both analyses was 48% and 58%, respectively, so while a slight improvement was seen, the results are poor and highly variable. These results demonstrate the difficulty in identifying deer using this method. While some individual deer possess highly distinctive face patterns and colouration, most in this population are hard to distinguish, either subjectively or based on their morphological measurements. The best results were achieved on the third analysis when the population was restricted to a small known sample size and the photographs were restricted to the highest quality and were taken on the same day (FRR=11%, FAR=0% and matching success=80%). This eliminated any seasonal and lighting differences and also reduced the photo error rate to 32%. The photographs used for this dataset were also taken concurrently such that the angles and background in the pictures do not differ by much, making matching much easier, particularly at the final subjective step.

Small populations are generally better suited to individual identification, as there is less chance of a duplication of measurements (Osterreider et al., 2015), particularly with methods that relies on photogrammetry. This is in contrast to pattern recognition algorithms that rely on unique 'fingerprints' such as spots and stripes, where duplication of patterns is much less likely, and large databases of photographs of large populations can be successfully analyzed (Bolger et al., 2012). While reasonable FRR/FAR rates were achieved for deer in the third analysis, the dataset may be too idealized to have produced misidentification rates that could be useful in real world applications. A truly successful individual identification method for deer may need to incorporate other morphological measurements (i.e., from additional photographs of the body and the side and/or the rear and tail patch). Identification of bobcats (*Lynx rufus*) and lynx (*Lynx canadensis*) often require full-body side views to accurately distinguish between these species: these are often available from camera trap data (Thornton, 2019). Pumas have been successfully identified from whole-body side view photographs from camera traps, by using scars, marks, colouration and spot patterns (Kelly et al., 2008); and coyotes have been distinguished based on side body views of their pelage patterns from camera trap photographs (Larrucea et al., 2007). Deer may possess these same identifying characteristics, either in combination with facial features or on their own. It should also be noted that the identification of deer using multiple photographs from different views may not require use of the program and may be achieved simply through subjective matching as was done with the above-noted methods and other facial recognition such as that of bottlenose dolphins (Genov et al., 2017).

Facial recognition software has advanced significantly in the last decade and even over the course of this study such that it is now been applied to several new species including lions, salmon and northern right whales (Kerr, 2015; Brueck, 2016; Daley, 2018). These methods use deep learning artificial intelligence, and in some cases crowd-sourcing, to achieve their results. They may prove applicable to deer; however, they are as-yet unpublished and may be unattainable for the average biologist due to the large computational requirements and high cost. In the case of salmon, the method was developed by a private company for use by fish farms, and in the case of right whales the program was developed by a private data science company. The development of the lion facial recognition initiative was projected to cost $65,000 (LINC, 2014). In the published literature, facial recognition has been used on other species such as elephant seals, tigers, and chimpanzees (Caiafa et al., 2005; Mason, 2016; Schofield et al., 2019). The high degree of accuracy (100%) achieved by the Eigenfaces method that was applied to elephant seals makes it a compelling method, but, it is highly sensitive to differences in the conditions in which the photographs are taken and thus often requires extensive pre-processing of the images and exclusion

of poorer quality images (Caiafa et al., 2005). In the case of lemurs and turtles, facial recognition is achieved using their hair or scale patterns (Schofield et al., 2008; Crouse et al., 2017). Deer have unique fur colours and patterns on their faces, but, they are not as overt as lemurs and therefore it is unclear whether they can be identified using the same type of feature extraction method.

A test of this photogrammetric method on male deer with antlers may also prove to be a feasible application, although year-to-year differences in antler size and shape would prohibit comparisons using photographs from sessions separated by >1 week or so. Merkle and Fortin (2013), however, suggest that changing traits could be accounted for by modifying the method using measurement-based transformations; antlers may be able to be incorporated in this way. Age and antler size of deer has been successfully estimated using morphometric ratios measured from photographs (Flinn, 2010) and could be similarly incorporated into this method to achieve individual identification.

A data simulation conducted by Merkle and Fortin (2013) demonstrated the impact of photo error on FRRs: a higher degree of photo error resulted in a higher FRR. A greater than 50% average photo error was considered high in the simulation. The impact of photo error on all of the analyses conducted is depicted in Figure 7. In general, as photo error increases, FRR/FAR also increase. The muskox analysis had a higher photo error (37%) compared to the low rates of FRR/FAR (0% and 3%) which was likely due to a high proportion of bad-angle photographs (~30%) in a small dataset (Figure 11). The badly-angled photographs were those in which the muskox was feeding and looking down, or at an angle of ~20 degrees. This increased the average photo error but because the dataset was small and the horn measurements were variable enough, they did not impede matching. Photo quality, lighting and camera angle have been shown to have a large impact on the accuracy of photo-identification methods across a wide range of studies (Kelly, 2001; Beekmans et al., 2005; Speed et al., 2007; Sherley at el, 2010). While the capture of high-quality photographs is the goal, broader application of this method for use with camera trap and real-world data may necessitate the use (or exclusion) of poor quality photographs, thus understanding misidentification rates is vital. Figure 10 depicts some of the photographs removed from the first deer analysis in an attempt to improve matching success. Ten photographs were removed from the first analysis to improve accuracy which resulted in the photo error dropping from 77% to 74% but did not significantly improve matching success, likely due to the inherent difficulty in identifying deer with this method.

5.2 Horns as a diagnostic trait

Attempting to use only facial measurements and not relying on horn measurements for sheep and goats resulted in lower success rates. The importance of the horn measurements was further emphasized by the results of the third goat analysis in which the horn length measurements were removed: the FRR rose from 11% to 28% and the FAR rose from 6% to 9% for the same set of photographs. The majority of the misidentifications from this analysis were program false rejections (11 of 17), therefore the horn length measurements were a crucial variable for successful program matching. Despite the apparent lack of variability in goat horns, they were an important morphological identifier for this species. The success of the muskox and sheep analyses, which relied most heavily on horn measurements, as compared to the deer analysis, also supports the conclusion that horns are a key characteristic for the identification of bovids with this method. Horns have been found to be effective identifiers of other bovid species such as long tailed gorals (*Nemorhaedus caudatus*; Zaumyslova & Bondarchuk, 2015). Horn morphology can also be used as a predictor of survival and reproductive success (Bergeron, 2012); thus the morphological measurements of horns used here may also be useful in understanding population and behavioural ecology. As mentioned above, antlers may also provide a useful and variable phenotypic trait for measurement. A standardized method of measurement that can be broadly applied to photographs of antlered species may allow for identification of individual male deer using photogrammetry. Another key component in developing this methodology for deer would be the development of a transformation to account for year-to-year changes in antler growth. A measurement-based transformation may also be required for species with horns, and would be particularly important in the identification of male sheep which can have much larger annual changes in horn growth than females. This presents an important next avenue of investigation for this method. Yoshikazi et al. (2009) have attempted to develop a method to incorporate evolving natural marks into closed population PMR models using simulations from a study on mark migration in larvae of northern two-lined salamanders (*Eurycea bislineata*). They found that misidentification errors due to evolving natural marks can cause population estimates to be significantly overestimated. However, misidentification errors are species and study specific and thus must be developed and applied in each case individually.

5.3 The impact of seasonality

One particularly interesting outcome of this study is the impact of seasonality on the performance of the matching program. When photographs of the same population of sheep were taken in the spring and in the fall and were combined into one dataset for the first analysis, the FRR was 11% and the FAR was 1%;

when the spring photographs were segregated in the second, the FRR was 4% and the FAR was 2%. In the case of sheep, this was likely due to the growth of fur around the base of their horns which affected where the horn base diameter measurement was taken. Additionally, the subjective choice made during the final step of the program was more difficult when comparing across seasons due to the change in colouration of the sheep's faces from winter fur growth. This difference can be seen in the photographs in Figure 9. Seasonal effects were also seen in the deer analyses. The first two deer analyses both contained photographs from October to November and a distinct difference can be seen between those taken in October and those taken in late November when the deer have grown, or started to grow, their winter coats. In the third deer analysis, the photographs are from one day in October and the FRR and FAR were much lower (11% and 0% vs. 22% and 12%). As noted above, the success of this analysis is likely due mostly to a small sample size and database as well as the photographs being highly idealized; however, some of the success may also be due to a lack of seasonal differences. The impact of seasonal differences on photo-identification has not been addressed in the literature and presents an important area for future investigation.

5.3 Observer Bias

Another avenue of investigation that has been addressed in the literature is the impact of observer bias on photo-identification. In the observer bias session that I conducted, there were significant differences between observers in matching choices. For deer, inter-rater agreement (or ICC) was 0.142 (95% CI=0.033-0.339) and for sheep it was 0.544 (95% CI=0.392-0.71). In general, ICC values less than 0.5 indicate poor agreement and values between 0.5-0.75 indicate moderate agreement (Koo & Li, 2016). Due to the large confidence interval (0.392-0.71) for sheep, this measure is less reliable so it can be said that there was significant observer bias for both deer and sheep, and poor inter-rater agreement for deer.

Additionally, the matching success rate for sheep averaged 85% while the matching success rate for deer was 26%. For a group of untrained observers, 85% success for sheep is a good result (compared to 96% matching success for an experienced observer on the same analysis), particularly considering the non-ideal conditions: the observers were not able to zoom into the pictures or individually examine them due to the nature of the session. The low rate of matching success for deer is also understandable considering the difficulty in identifying deer: 35% matching success was achieved for the same analysis by an experienced observer. Observer bias in photo-identification studies has been previously addressed in the literature. In a camera trap study on lowland tapirs (*Tapirus terrestris*), 14 observers

were tested on their ability to correctly identify 8 known tapirs from 55 photographs and were found to range from over-estimating the population by 75% to underestimating it by 50% (Oliviera-Santos et al., 2009). In a similar camera trap study on pumas (*Puma concolor*), however, average agreement between pairs of observers was found to be 79% (Kelly et al., 2008). Rates of agreement between observers is largely dependent on the species and method. Mendoza et al. (2011) designed an online classification tool to improve individual identification of animals from camera trap data. Application of their method resulted in improved agreement from 0.28 to 0.84 (where perfect agreement=1) between 2 observers on the identification of bobcats (*Lynx rufus*; Mendoza et al., 2011). In a facial recognition study of loggerhead sea turtles, an expert observer obtained matching success of 99% while naïve observers obtained a matching success rate of 71%, but with training and repetition this was improved to 87% (Schofield et al., 2008).

In addition, the use of photographic matching software appears to reduce observer error rates considerably when compared to unaided visual matching (Cruikshank & Schmidt, 2017). For a relatively easy to identify species like yellow-bellied toads (*Bombina variegata*), bobcats, and sheep, observer bias can be low and can be managed if it is understood and acknowledged (Cruikshank & Schmidt, 2017; Mendoza et al., 2011; whereas with difficult to distinguish species like deer, pumas, and tapirs, observer bias may be quite high and difficult to overcome (Oliviera-Santos et al., 2009; Kelly et al., 2008). Training and experience should be emphasized in the use of this method.

5.8 User choice

The final subjective stage of the method, in which the observer chooses the correct match, was a crucial step in this method. While observer bias may affect the outcome of the program and should be taken into consideration, the calculated automated matching success rates point to the importance of subjective user choice in the final step of the program. A significant improvement in the matching success of the program was found when comparing the automated matching success with the program matching success. For certain species, there is potential for calculating a correction factor that could be used to adjust population size estimates. A correction factor could be determined by running repeated analyses or data simulations to determine whether a consistent difference exists between the program and automated matching success. This would make it possible to eliminate the final subjective choice step which can be both time consuming and potentially biased. This is particularly true for species for which the method works well (i.e., for muskox and sheep), in which the correct match appears in the top ranked spot more frequently. Additionally, these results point to the inverse possibility that for such

small population sizes and datasets, the use of the program may not be necessary when user choice is required and also more accurate. Due to the time-consuming nature of pre-processing the photographs and running the program that may be true for this data, however, when assessing larger populations with larger datasets the program retains its utility.

6.0 Conclusions

I tested Merkle and Fortin's (2013) photogrammetric approach to facial recognition on 4 species of ungulates which included muskox, sheep, goats and deer. The main finding of this study was that the success of this approach to individual recognition of ungulates was species specific. While the method I used was accurate for sheep and muskox, moderate results were achieved for goats and generally poor results were achieved for deer. In addition, seasonal effects were found when photographs captured in spring and fall were combined in the same dataset. Increasing the number of measurements taken increased success rates, however, the horns proved to be the most important predictor of success: they provide a highly variable, unique and easily measureable phenotypic trait for identification. Consideration of potential observer bias is important in the use of the program and training should be provided for naïve observers. Subjective user choice was also an important component of the method despite the fact that it is time-consuming and introduces bias.

The photographic identification method tested and expanded here provides a novel and non-invasive tool for population monitoring. The technique is applicable to the identification of unique individuals of the three bovid species tested and may prove useful in mark-recapture studies. Non-invasive population monitoring is an important tool in sensitive and inaccessible areas: this method may allow for expanded monitoring of species that have been previously difficult or expensive to monitor. Further investigation is required to determine the limits of the method, particularly for sheep and muskox, particularly on the effects of seasonality on the success of the method, and also on the applicability of the method from year-to-year.

Next steps include testing the method on a larger and wild population of sheep and muskox, and attempting to use camera trap data for the purposes of identification. Application of the method on a known wild population in a mark-recapture study would allow real-world misidentification rates to be determined and incorporated. The assumptions of mark-recapture would be upheld in a short-term study, but for longer-term studies quantification of the stability of the traits (i.e., horns) would be required.

Determining whether camera traps can capture useful and high quality photographs of sheep and muskox for use in individual identification would also be important for the further expansion of the applicability of the method. The application of this method to deer and goats may require more creativity, but with the advent of facial recognition software for humans and its expansion to some species of primates (Loos & Ernst, 2013), the development of facial recognition software for other mammal species is the next logical step. This study provides a basis for the development of such software by establishing the facial measurements for which there is adequate variability on which to build such an algorithm. It also provides a basis for applying PMR to northern ungulate species by establishing misidentification rates for 4 species, as well as some of the key factors to consider in its application, including observer bias and seasonal effects.

7.0 References

Abramoff MD, Magalhaes PJ, & Ram SJ. 2004. Image processing with ImageJ. Biophotonics International 11:36–42.

Alonso RS, McClintock BT, Lyren LM, Boydston EE, Crooks KR. 2015. Mark-recapture and mark-resight methods for estimating abundance with remote cameras: a carnivore case study. PLoS One. 10:e0123032.

Anderson CJR, Lobo ND, Roth JD and Waterman JM. 2010. Computer-aided photo-identification system with an application to polar bears based on whisker spot patterns. Journal of Mammalogy 91:1350–1359.

Araabi B, Kehtarnavaz N, McKinney T, Hillman G, Würsig B. 2000. A string matching computer-assisted system for dolphin photoidentification. Annals of Biomedical Engineering. 28:1269-79.

Ardovini A, Cinque L, Sangineto E. 2008. Identifying elephant photos by multi-curve matching. Pattern Recognition. 41:1867-77.

Arnemo JE, Ahlqvist P, Andersen R, Berntsen F, Ericsson G, Odden J, Brunberg S, Segerström P, Swenson JE. 2006. Risk of capture-related mortality in large free-ranging mammals: experiences from Scandinavia. Wildlife Biology. 12:109-113.

Arzoumanian Z, Holmberg J, Norman B. 2005. An astronomical pattern-matching algorithm for computer-aided identification of whale sharks *Rhincodon typus*. Journal of Applied Ecology. 42:999-1011.

Ballard WB & Tobey RW. 1981. Decreased calf production of moose immobilized with anectine administered from helicopter. Wildlife Society Bulletin. 9: 207-209.

Becciolini V, Lanini F & Ponzetti MP. 2019. Impact of capture and chemical immobilization on the spatial behaviour of red deer (*Cervus elaphus*) hinds. Wildlife Biology. Downloaded from https://bioone.org/journals/Wildlife-Biology on 1 August 2019.

Beekmans, BWPM, Whitehead H, Huele R, Steiner L & Steenbeck AG. 2008. Comparison of two computer-assisted photo-identification methods applied to sperm whales (*Physeter macrocephalus*) Aquatic Mammals. 31:243-247.

Bell AM, Hankison SJ, and Laskowski KL. 2009. The repeatability of behaviour: a meta-analysis. Animal Behaviour. 77:771-783.

Berger J. 2012. Estimation of body-size traits by photogrammetry in large mammals to inform conservation. Conservation Biology. 26:769-777.

Bergeron P. 2007. Parallel lasers for remote measurements of morphological traits. Journal of Wildlife Management. 71:289-292.

Bolger DT, Morrison TA, Vance B, Lee D, Farid H. 2012. A computer-assisted system for photographic mark-recapture analysis. Methods in Ecology and Evolution. 3:813-822.

Brueck, H. Jan. 19, 2016. A surprising tool for saving the whales: facial recognition software. Fortune. Available from https://fortune.com/2016/01/19/facial-recognition-whales/. Accessed on September 6, 2019.

Burton C, Neilson E, Moreira D, Ladle A, Steenweg R, Fisher JT, Bayne E & Boutin S. 2015. Wildlife camera trapping: a review and recommendations for linking surveys to ecological processes. Journal of Applied Ecology. 52:675–685.

Byers T. 1999. Perspectives of Aboriginal Peoples on wildlife research. Wildlife Society Bulletin. 27: 671-675.

Caiafa CF, Proto AN, Vergani D, Stanganelli Z. 2005. Development of individual recognition of female southern elephant seals, *Mirounga leonina*, from Punta Norte Península Valdés, applying principal components analysis. Journal of Biogeography. 32:1257-1266.

Cattet M, Boulanger J, Stenhouse G, Powell RA, Reynolds-Hogland MJ. 2008. An evaluation of long-term capture effects in ursids: Implications for wildlife welfare and research. Journal of Mammalogy. 89:973-990.

Côté SD, Festa-Bianchet M, Fournier F. 1998. Life-history effects of chemical immobilization and radiocollars on mountain goats. The Journal of Wildlife Management. 62:745-752.

Côté SD, Festa-Bianchet M, Smith KG. 1998. Horn growth in mountain goats (*Oreamnos americanus*). Journal of Mammalogy. 79:406-414.

Crall JP, Stewart CV, Berger-Wolf TY, Rubenstein DI, Sundaresan SR. 2013. HotSpotter - Patterned Species Instance Recognition. Available from http://cs.rpi.edu/hotspotter/crall-hotspotter-wacv-2013.pdf. Accessed on March 31, 2019.

Crouse D, Jacobs RL, Richardson Z, Klum S, Jain A, Baden AL & Tecot SR. 2017. LemurFaceID: a face recognition system to facilitate individual identification of lemurs. BMC Zoology. 2:2.

Cruikshank SS & Schmidt BR. 2017. Error rates and variation between observers are reduced with the use of photographic matching software for capture-recapture studies. Amphibia-Reptilia. 38: 315-325.

Cutler TL, Swann DE. 1999. Using remote photography in wildlife ecology: a review. Wildlife Society Bulletin (1973-2006). 27:571-581.

Daley J. Oct. 9, 2018. How fish farms can use facial recognition to survey sick salmon. Smithsonian Magazine. Available from https://www.smithsonianmag.com/smart-news/facial-recognition-will-be-used-monitor-fish-faces-180970493/. Accessed on September 6, 2019.

Darimont CT, Reimchen TE, Bryan HM, & Paquet PC. 2008. Faecal-centric approaches to wildlife ecology and conservation; methods, data and ethics. Wildlife Biology in Practice. 4:73-87.

Dechen-Quinn AC, Williams DM, Porter WF, Fitzgerald SD, Hynes K. 2014. Effects of capture related injury on post-capture movement of white-tailed deer. Journal of Wildlife Diseases. 50:250-258.

Delehanty B & Boonstra R. 2009. Impact of live trapping on stress profiles of Richardson's ground squirrel (*Spermophilus richardsonii*). General and Comparative Endocrinology. 160:176-182.

DelGiudice GD, Severud WJ, Obermoller TR Wright RG, Enright TA. 2015. Monitoring movement behavior enhances recognition and understanding of capture-induced abandonment of moose neonates. Journal of Mammalogy. 96:1005-1016.

Dertien JS, Doherty PF, Bagley CF, Haddix JA, Brinkman AR, Neipert ES. 2017. Evaluating dall's sheep habitat use via camera traps. The Journal of Wildlife Management. 81:1457-1467.

Drechsler A, Helling T, Steinfartz S. 2015. Genetic fingerprinting proves cross-correlated automatic photo-identification of individuals as highly efficient in large capture–mark–recapture studies. Ecology and Evolution. 5:141-51.

Flinn JJ. 2010. Accuracy of estimating age and antler size of photographed deer. MSc Thesis. ProQuest Dissertations Publishing.

Forcada J, Aguilar A. 2000. Use of photographic identification in capture-recapture studies of mediterranean monk seals. Marine Mammal Science. 16:767-793.

Foster G, Krijger H, Bangay S. 2006. Zebra fingerprints: Towards a computer-aided identification system for individual zebra. African Journal of Ecology. 45:225-227.

Gamble L, Ravela S, McGarigal K. 2008. Multi-scale features for identifying individuals in large biological databases: An application of pattern recognition technology to the marbled salamander (*Ambystoma opacum)*. Journal of Applied Ecology. 45:170-180.

Gee KL, Webb SL, Holman JH. 2014. Accuracy and implications of visually estimating age of male white-tailed deer using physical characteristics from photographs. Wildlife Society Bulletin. 38:96-102.

Genov T, Centrih T, Wright AJ & Wu GM. 2017. Novel method for identifying individual cetaceans using facial features and symmetry: a test case using dolphins. Marine Mammal Science. 34:514-528.

Gope C, Kehtarnavaz N, Hillman G, Würsig B. 2005. An affine invariant curve matching method for photo-identification of marine mammals. Pattern Recognition. 38:125-32.

Goswami VR, Madhusudan MD, Karanth KU. 2007. Application of photographic capture–recapture modelling to estimate demographic parameters for male asian elephants. Animal Conservation. 10:391-9.

Hayes JP and Jenkins SH. 1997. Individual variation in mammals. Journal of Mammalogy. 78:274-293.

Hiby L & Lovell P. 1990. Computer aided matching of natural markings: a prototype system for grey seals. Reports of the International Whaling Commission (sp. Issue 12):57-61.

Hiby L, Paterson WD, Redman P, Watkins J, Twiss SD, Pomeroy P, Freckleton R. 2013. Analysis of photo-id data allowing for missed matches and individuals identified from opposite sides. Methods in Ecology and Evolution. 4:252-259.

Higashide D, Miura S, Miguchi H. 2012. Are chest marks unique to asiatic black bear individuals? Journal of Zoology. 288:199-206.

Hillman GR, Würsig B, Gailey GA, Kehtarnavaz N, Drobyshevsky A, Araabi BN, Tagare HD, Weller DW. 2003. Computer-assisted photo-identification of individual marine vertebrates: a multi-species system. Aquatic Mammals. 29:117–123.

Hopkins M & Milton K. 2016. Adverse effects of ball-chain radio-collars on female mantled howlers (*Alouatta palliata*) in panama. International Journal of Primatology. 37:213-224.

Huffard CH, Caldwell RL, DeLoach N, Gentry DW, Humann P, MacDonald B, Moore B, Ross R, Uno T, Wong S. 2008. Individually unique body color patterns in octopus (*Wunderpus photogenicus*) allow for photoidentification. PLoS One. 3:e3732.

Jacques CN, Jenks JA, Depemo CS, Sievers JD, Grovenburg JW, Brinkman TJ, Swanson CC, and Stillings BA. 2009. Evaluating ungulate mortality associated with helicopter net-gun captures in the Northern Great Plains. Journal of Wildlife Management. 73:1282–1291.

Jung TS, Konkolics SM, Kukka PM, Majchrzak YN, Menzies AK, Oakley MP, Peers MJL & Studd EK. 2019. Short-term effect of helicopter-based capture on movements of a social ungulate. Journal of Wildlife Management. 83:830-837.

Karanth KU, Nichols JD. 1998. Estimation of tiger densities in india using photographic captures and recaptures. Ecology. 79:2852-2862.

Kerr M. Jul. 1, 2015. Lion facial recognition debuts in Africa. Scientific American. Available from https://www.scientificamerican.com/article/lion-facial-recognition-debuts-in-africa/. Accessed on September 6, 2019.

Krausman PR, Bleich VC, Cain III JW, Stephenson TR, DeYoung DW, McGrath PW, Swift PK, Pierce BM, Jansen BD. 2004. Neck lesions in ungulates from collars incorporating satellite technology. Wildlife Society Bulletin. 32:987-991.

Kelly MJ, Noss AJ, Di Bitetti MS, Maffei L, Arispe RL, Paviolo A, De Angelo CD & Di Blanco YE. 2008. Estimating Puma densities from camera trapping across three study sites: Bolivia, Argentina and Belize. Journal of Mammalogy. 89:408–418.

Kelly MJ. 2001. Computer-aided photograph matching in studies using individual identification: An example from serengeti cheetahs. Journal of Mammalogy. 82:440-449.

Koo TK & Li MY. 2016. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. Journal of Chiropractic Medicine. 15:155-163.

Larrucea ES, Brussard PF, Jaeger MM, Barrett RH. 2007. Cameras, coyotes, and the assumption of equal detectability. Journal of Wildlife Management. 71:1682–1689.

Larsen DG & Gauthier DA. 1989. Effects of capturing pregnant moose and calves on calf survivorship. Journal of Wildlife Management. 53:564-567.

LINC. 2014. Lion Identification Network of Collaborators. Available from http://lionguardians.org/wp-content/uploads/2014/11/Make-The-LINC.pdf. Accessed on September 6th, 2019.

Livezey KB. 1990. Toward the reduction of marking-induced abandonment of newborn ungulates. Wildlife Society Bulletin. 18:193-203.

Loos A & Ernst A. 2013. An automated chimpanzee identification system using face detection and recognition. EURASIP Journal on Image and Video Processing. 49.

Lubow BC & Ransom JL. 2009. Validating aerial photographic mark–recapture for naturally marked feral horses. Journal of Wildlife Management. 73:1420-1429.

Magoun AJ, Long CD, Schwartz MK, Pilgrim KL, Lowell RE, Valkenburg P. 2011. Integrating motion-detection cameras and hair snags for wolverine identification. Journal of Wildlife Management. 75:731-739.

Mason AD. 2016. Monitoring individual animals through a collaborative crowdsourcing and citizen science platform. Thesis Dissertation. University of Surrey, UK. ProQuest Dissertations Publishing.

Mendoza E, Martineau PR, Brenner E, Dirzo R. 2011. A novel method to improve individual animal identification based on camera-trapping data. Journal of Wildlife Management. 75:973-979.

Merkle JA, Fortin D. 2014. Likelihood-based photograph identification: Application with photographs of free-ranging bison. Wildlife Society Bulletin. 38:196-204.

Morrison TA, Bolger DT. 2012. Wet season range fidelity in a tropical migratory ungulate. Journal of Animal Ecology. 81:543-552.

Northrup JM, Anderson CR, and Wittemyer G. 2014. Effects of helicopter capture and handling on movement behavior of mule deer. Journal of Wildlife Management. 78:731–738.

Ngoprasert D, Reed DH, Steinmetz R, and Gale GA. 2012. Density estimation of Asian bears using photographic capture–recapture sampling based on chest marks. Ursus 23:117–133.

Oliveira-Santos LGR, Zucco CA, Antunes PC & Crawshaw Jr. PG. 2009. Is it possible to individually identify mammals with no natural markings using camera traps? A controlled case-study with lowland tapirs. Mammalian Biology. 75:375-378

Osterrieder SK, Kent CS, Anderson CJR, Parnum IM, Robinson RW. 2015. Whisker spot patterns: a noninvasive method of individual identification of Australian sea lions (*Neophoca cinerea*). Journal of Mammalogy. 96:988-997.

Otis DL, Burnham KP, White GC & Anderson DR. 1978. Statistical inference from capture data on closed animal populations. Wildlife Monographs. 62:3-135.

Powell RA, Proulx G. 2003. Trapping and marking terrestrial mammals for research: Integrating ethics, performance criteria, techniques, and common sense. ILAR Journal. 44:259-276.

Rachlow JL, Peter RM, Shipley LA, and Johnson TR. 2014. Sublethal effects of capture and collaring on wildlife: experimental and field evidence. Wildlife Society Bulletin. 38:458–465.

Rode KD, Pagano AM, Bromaghin JF, Atwood TC, Durner GM, Simac KS, and Amstrup SC. 2014. Effects of capturing and collaring on polar bears: findings from long-term research on the southern Beaufort Sea population. Wildlife Research. 41:311–322.

Rockhill AP, Sollman R, Powell RA, DePerno CS. 2016. A comparison of survey techniques for medium- to large-sized mammals in forested wetlands. Southeastern Naturalist. 15:175-187.

Sacchi R, Scali S, Pellitteri-Rosa D, Pupin F, Gentilli A, Tettamanti S, Cavigioli L, Racina L, Maiocchi V, Galeotti P, and Fasola M. 2010. Photographic identification in reptiles: a matter of scales. AmphibiaReptilia 31:489–502.

Schmidt JH, Rattenbury KL, Robison HL, Gorn TS, Shults BS. 2017. Using non-invasive mark-resight and sign occupancy surveys to monitor low-density brown bear populations across large landscapes. Biological Conservation. 207:47-54.

Schofield D, Nagrani A, Zisserman A, Hayashi M, Matsuzawa T, Biro D & Carvalho S. 2019. Chimpanzee face recognition from videos in the wild using deep learning. Science Advances. 5: eaaw0736.

Schofield GG, Katselidis KA, Dimopoulos P & Panti JD. 2008. Investigating the viability of photo-identification as an objective tool to study endangered sea turtle populations. Journal of Experimental Marine Biology and Ecology. 360:103–108.

Schwarz CJ & Seber GAF. 1999. Estimating Animal Abundance: Review III. Statistical Science. 14:427-456.

Seber GAF. 1982. The Estimation of Animal Abundance and Related Parameters. 2nd ed. Arnold, London.

Sherley R, Burghardt T, Barham P, Campbell N, Cuthill I. 2010. Spotting the difference: Towards fully-automated population monitoring of african penguins (*Spheniscus demersus)*. Endangered Species Research. 11:101-111.

Soria-Díaz L, Monroy-Vilchis O. 2015. Monitoring population density and activity pattern of white-tailed deer (odocoileus virginianus) in central mexico, using camera trapping. Mammalia. 79:43-50.

Speed CW, Meekan MG, Bradshaw CJA. 2007. Spot the match - wildlife photo-identification using information theory. Frontiers in Zoology. 4:2.

Swenson JE, Wallin K, Ericsson G, Cederlund G, Sandegren F. 1999. Effects of ear-tagging with radiotransmitters on survival of moose calves. Journal of Wildlife Management. 63:354-358.

Tancredi A, Auger-Méthé M, Marcoux M, Liseo B. 2013. Accounting for matching uncertainty in two stage capture–recapture experiments using photographic measurements of natural marks. Environmental and Ecological Statistics. 20:647-665.

Thornton DH. 2019. Reassessing the success of experts and nonexperts at correctly differentiating between closely related species from camera trap images: a reply to Gooliaff and Hodges. Ecology and Evolution. 9:6172–6175.

Turk & Pentland, 1991. Face recognition using eigenfaces. IEEE Computer Society Conference on Computer Vision and Pattern Recognition Computer Vision and Pattern Recognition, p. 586-591.

Walker KA, Trites AW, Haulena M, Weary DM. 2012. A review of the effects of different marking and tagging techniques on marine mammals. Wildlife Research. 39:15-30.

White M, Knowlton FF & Glazener WC. 1972. Effects of dam-newborn fawn behavior on capture and mortality. Journal of Wildlife Management. 36:897-906.

Yoshikazi J, Pollock KH, Brownie C & Webster RA. 2009. Modeling misidentification errors in capture–recapture studies using photographic identification of evolving marks. Ecology. 90:3–9.

Zaumyslova OY & Bondarchuk SN. 2015. The use of camera traps for monitoring the population of long-tailed gorals. Achievements in the Life Sciences 9:15 –21.

Appendix 1: Tables and Figures

Table 1: Summary of photographic identification studies

Species	Identification Characteristic	Method	Sample Size	Reliability	Image Processing Program	Citation
African Penguins (*Spheniscus demersus*)	Ventral spots	Computer-assisted photograph matching	114 (1000 images)	>90% genuine acceptance rate; FAR=0.03%	AnimalID	Sherley *et al*. (2010)
Asian elephant (*Elephus maximus*)	Tusks, scars and other morphological characteristics	Photographic CR	78 (135 sightings)	FAR=0.007	n/a	Goswami *et al*. (2007)
Asiatic black bear (*Ursus thibetanus*)	Chest marks	Image analysis	52 (735 images)	FAR=0.00075	ImageJ	Higashide *et al*. (2012)
Australian sea lions (*Neophoca cinerea*)	Whisker spot patterns	Pattern recognition algorithm using Chamfer distance transform	53 (608 images)	99%-88% accurate for simulated population of 50-1000	n/a	Osterreider *et al*. (2015)
Cheetah (*Acinonyx jubatus*)	Spot patterns	3D computer-aided matching	10,000 images	FRR=6.5%; FAR=2.5%	see Hiby & Lovell (1990)	Kelly (2001)
Dolphins (*Tursiops truncatus*)	Dorsal fin notch patterns	String matching	164 (624 images)	n/a	n/a	Araabi *et al*. (2000)
Dusky dolphins (*Lagenorhynchus obscurus*), spinner dolphins (*Stenella longirostris*), long-finned pilot whales (*Globicephala melas*), white sharks (*Carcharodon carcharias*) and sperm whale (*Physeter macrocephalus*)	Fin shape (and fluke shape in the case of sperm whales)	Computer assisted photograph matching via curve matching and string matching	30-44 (23-127 additional images)	Correct identification on first suggested match in 50% of the images; 75% of the time in the first 3-4 images	FinScan	Hillman *et al*. (2003)
Elephants (*Loxodonta* spp.)	Ear edge nick patterns	Multi-curve matching	60 (111) tested; 268 (332 images) total	83% true positives	n/a	Ardovini *et al*. (2008)
Giraffes (*Giraffa camelopardalis*), jaguars (*Panthera onca*), lionfish (*Pterois* sp.), plains zebras (*Equus quagga*) and Grevy's zebras (*Equus grevyi*)	Stripe and spot patterns	Computer assisted photograph matching via one-vs-one and one-vs-many matching	86 (824) plains zebras; 592 (1047) Grevy's zebras; 21 (45) jaguar; 15 (45) giraffes; 5 (13) lionfish	100% for giraffes, lionfish and jaguar; 95-98% for Grevy's zebras; 99% for plains zebras	HotSpotter	Crall *et al*. (2013)
Great crested newt (*Triturus cristatus*)	Ventral spot patterns	Pattern extraction using an automated cross-correlation algorithm	100 images for FRR; 1648 images total with 206 recaptures	78% of recaptures ID'd; FRR=2%; FAR=0%	AMPHIDENT	Drechsler *et al.* (2015)
Grey seal (*Halichoerus grypus*)	Pelage pattern	Multi-biometric ID and PMR	156 (155 pairs of images)	33% FRR; 14% FRR with only 'good' images	ExtractCompare	Hiby *et al*. (2013)
Harbor seals (*Phoca vitulina richardii*)	Pelage, scars and flipper tags	Computer-assisted photograph matching	182 (772 images)	1.8% misidentification error rate	ExtractCompare	Hastings *at al*. (2008)
Marbled salamander (*Ambystoma opacum*)	Dorsal patterns	Pattern matching algorithm	101 (1008 images); 447 total	95% success rate	n/a	Gamble *et al*. (2008)
Masai giraffe (*Giraffa camelopardalis tippelskirchi*)	Flank pattern (reticulated polygons)	Pattern extraction and comparison using Scale Invariant Feature Transform (SIFT)	50 (100 images) to test method; 568 (1026 images) to estimate population	FRR=0.007, FAR=0.000	Wild-ID	Bolger *et al*. (2012)

Table 1 (cont.): Summary of photographic identification studies

Species	Identification Characteristic	Method	Sample Size	Reliability	Image Processing Program	Citation
Mediterranean monk seals (*Monachus monachus*)	Marks, scars and pelage patterns	Observer matching	14 (42 images) for validation; 228 total	FAR=0.043		Forcada *et al*. (2000)
Octopus (*Wunderpus photogenicus*)	Dorsal mantle markings	Untrained observer matching	15 (30 images)	FRR=4.5%; FAR=18%;	Adobe Illustrator	Huffard *et al*. (2008)
Polar bears (*Ursus arctos*)	Whisker spot patterns	Computer-aided pattern matching algorithm (Chamfer distance)	57 (>1000 images)	80% true positives, 10% false positives	n/a	Anderson *et al*. (2010)
Puma (*Puma concolor*)	Marks, scars, coloring, etc	Observer matching	26-30 (214 images)	79.3% agreement between investigators across all sites	n/a	Kelly *et al*. (2008)
Sea lion (*Eumetopias jubatus*), grey whale (*Eschrichtius robustus*), dolphin	Fin, flipper and fluke edges	Affine invariant curve matching	37 (92 images); 37 (95 images); 164 (624 images)	Error rate~0.25	CurveMatch	Gope *et al*. (2005)
Whale shark (*Rhincodon typus*)	Spot patterns	Astronomical pattern matching algorithm	27 (~450 images) (from ECOOCEAN database)	90% success rate	n/a	Arzoumanian *et al.* (2005)
Whale shark (*Rhincodon typus*)	Spot patterns, scars and marks	Observer matching	159 (221 images)	95% confidence interval for population estimate	n/a	Meekan *et al*. (2006)
Whale sharks (*Rhincodon typus*)	Spot patterns	Computer-assisted photograph matching	50 (100 images)	93% match rate	I^3S	Speed *et al*. (2007)
White tailed deer (*Odocoileus virginianus*)	Fur (scars and spots), neck size, antler or fawn presence or absence	Observer matching	4-11 (94 images)	97% agreement between independent experts	n/a	Soria-Diaz & Monroy-Vilchis (2015)
Wildebeest (*Connochaetes taurinus*)	Shoulder stripe patterns	Computer-assisted photograph matching	198 (matching pairs of images)	FAR= 0.00081; FRR=0.06-0.08	Wild-ID (for females) and ExtractCompare (for males)	Morrison & Bolger (2012)
Wolverine (*Gulo gulo*)	Ventral patterns	Observer matching	18-21	100% agreement between photo ID and DNA	n/a	Magoun *et al*. (2011)
Zebra (*Equus burchelli*)	Stripe patterns	Image processing and fingerprint ID	6 out of 50 images for validation	79.8% ± 12.5% matching success	ImageJ plugin	Foster *et al*. (2006)

Table 2: Summary of facial recognition-based photographic identification studies

Species	Identification Characteristic	Method	Sample Size	Reliability	Image Processing Program	Citation
Bison (*Bison bison*)	Facial and horn measurements	Computer-assisted photograph matching	33 (91 images)	FRR=0.055	MatchImage	Merkle & Fortin (2013)
Bottlenose Dolphins (*Tursiops truncatus*)	Facial features	Observer matching	20 (40 images)	87.5% matching success	n/a	Genov et al. (2017)
Chimpanzees (*Pan troglodytes*)	Facial features	Facial identification	24-71 (2617-3905 images)	FRR=12.5-17.7%; FAR= 3.5-13.5%	n/a	Loos & Ernst (2013)
Elephant seals (*Mirounga leonina*)	Faces	Eigenfaces method	56 (96 images)	100%	Eigenfaces; IDL5.5	Caiafa *et al.* (2005)
Loggerhead sea turtles (*Caretta caretta*)	Facial scales	Identification Tree	50 (200 images)	99% matching success	n/a	Schofield *et al.* (2008)
Red-bellied Lemurs (*Eulemur rubriventer*)	Facial features	Feature extraction	80 (462 images)	98.7%±1.8% accuracy	LemurFaceID	Crouse *et al.* (2017)

Figure 1: Nine measurements (M1-M9) used for muskox analysis. Measurements were chosen based on those used by Merkle & Fortin (2013). See Table 3 for descriptions.

Table 3: Descriptions of nine measurements (M1-M9) taken in the program ImageJ and used for the muskox analysis. Measurements were modelled on those used by Merkle & Fortin (2013) for bison.

M1	Right horn tip to left horn tip
M2	Outermost point of right horn to outermost point of left horn
M3	Diameter of right horn at M2
M4	Diameter of left horn at M2
M5	Diameter of right horn where horn meets face
M6	Diameter of left horn where horn meets face
M7	Length of right horn from M2 to M5
M8	Length of left horn from M2 to M6
M9	Inner corner of right eye to inner corner of left eye

Table 4: Standard deviations and weights used to calculate similarity scores for muskox analysis in MatchImage. Standard deviations (photograph error) were calculated by measuring 4-5 photographs each of 5 different animals and then a mean standard deviation (SD) of each of the ratios was calculated. Weights (measurer error) were calculated by measuring the same photograph 5 times and repeating this for 10 individuals. The mean SD was then taken for each ratio and used to calculate the weights. The weights are multiplicative such that the smallest mean SD was weighted as 1 with the rest less than 1, based on their respective sizes.

Ratio	Measurement	SD	Weight
r1	M1/M9	0.101936	0.9695283
r2	M2/M9	0.066821	0.9625129
r3	M3/M9	0.010979	0.9935613
r4	M4/M9	0.015368	0.9925231
r5	M5/M9	0.019418	0.9917574
r6	M6/M9	0.017314	0.9878246
r7	M7/M9	0.049721	0.96089
r8	M8/M9	0.063294	0.9562241
r9	(M2-M1)/M1	0.017212	1

Figure 2: Twenty deer measurements (M1-M20) used for all 3 deer analyses (set 1, 2 & 3). Measurements were chosen based on those used in human facial recognition software, the traits available for measure in front facing photographs and those used by Merkle & Fortin (2013). See Table 5 for descriptions.

Table 5: Descriptions of 20 measurements (M1-M20) taken in the program ImageJ and used for all 3 deer analyses (Deer set 1, 2, & 3). Measurements were chosen based on those used in human facial recognition software, the traits available for measure in front facing photographs and those used by Merkle & Fortin (2013).

M1	Inner corner of right eye to inner corner of left eye
M2	Outermost point of right eye to outermost point of left eye
M3	Outside of right nostril to outside of left nostril at widest point
M4	Width of septum at narrowest point
M5	Width of base of right ear at narrowest point
M6	Width of base of left ear at narrowest point
M7	Length of right ear from tip to midpoint of M5
M8	Length of left ear from tip to midpoint of M6
M9	Left end of M5 to right end of M6
M10	Midpoint of M1 to midpoint M3
M11	Midpoint of M9 to midpoint of M2
M12	Bottom of M5 to bottom of M6
M13	Left end of M2 to left end of M3
M14	Right end of M2 to right end of M3
M15	Left end of M12 to left of M2
M16	Right end of M12 to right end of M2
M17	Left end of M1 to left end of M4
M18	Right end of M1 to right of M4
M19	Left end of M12 to right end of M1
M20	Right end of M12 to left end M1

Table 6: Standard deviations and weights used to calculate similarity scores for all 3 deer analyses in MatchImage. Standard deviations (photograph error) were calculated by measuring 4-5 photographs each of 5 different animals and then a mean standard deviation (SD) of each of the ratios was calculated. Weights (measurer error) were calculated by measuring the same photograph 5 times and repeating this for 10 individuals. The mean SD was then taken for each ratio and used to calculate the weights. The weights are multiplicative such that the smallest mean SD was weighted as 1 with the rest less than 1, based on their respective sizes.

Ratio	Measurement	SD	Weight
r1	M3-M4/M4	0.198	0.068
r2	M2/M1	0.024	1.000
r3	M3/M1	0.020	0.368
r4	M4/M1	0.019	0.200
r5	M5/M1	0.030	0.423
r6	M6/M1	0.025	0.516
r7	M7/M1	0.035	0.870
r8	M8/M1	0.038	0.836
r9	M9/M1	0.029	0.666
r10	M10/M1	0.028	0.413
r11	M11/M1	0.023	0.620
r12	M12/M1	0.054	0.629
r13	M13/M1	0.027	0.602
r14	M14/M1	0.026	0.591
r15	M15/M1	0.025	0.284
r16	M16/M1	0.030	0.215
r17	M17/M1	0.032	0.477
r18	M18/M1	0.028	0.509
r19	M19/M1	0.037	0.664
r20	M20/M1	0.036	0.725

Figure 3: Ten measurements (M1-M10) used for the sheep analysis 1 &2. The 10 measurements were modelled on those used by Merkle & Fortin (2013) for bison. See Table 7 for descriptions.

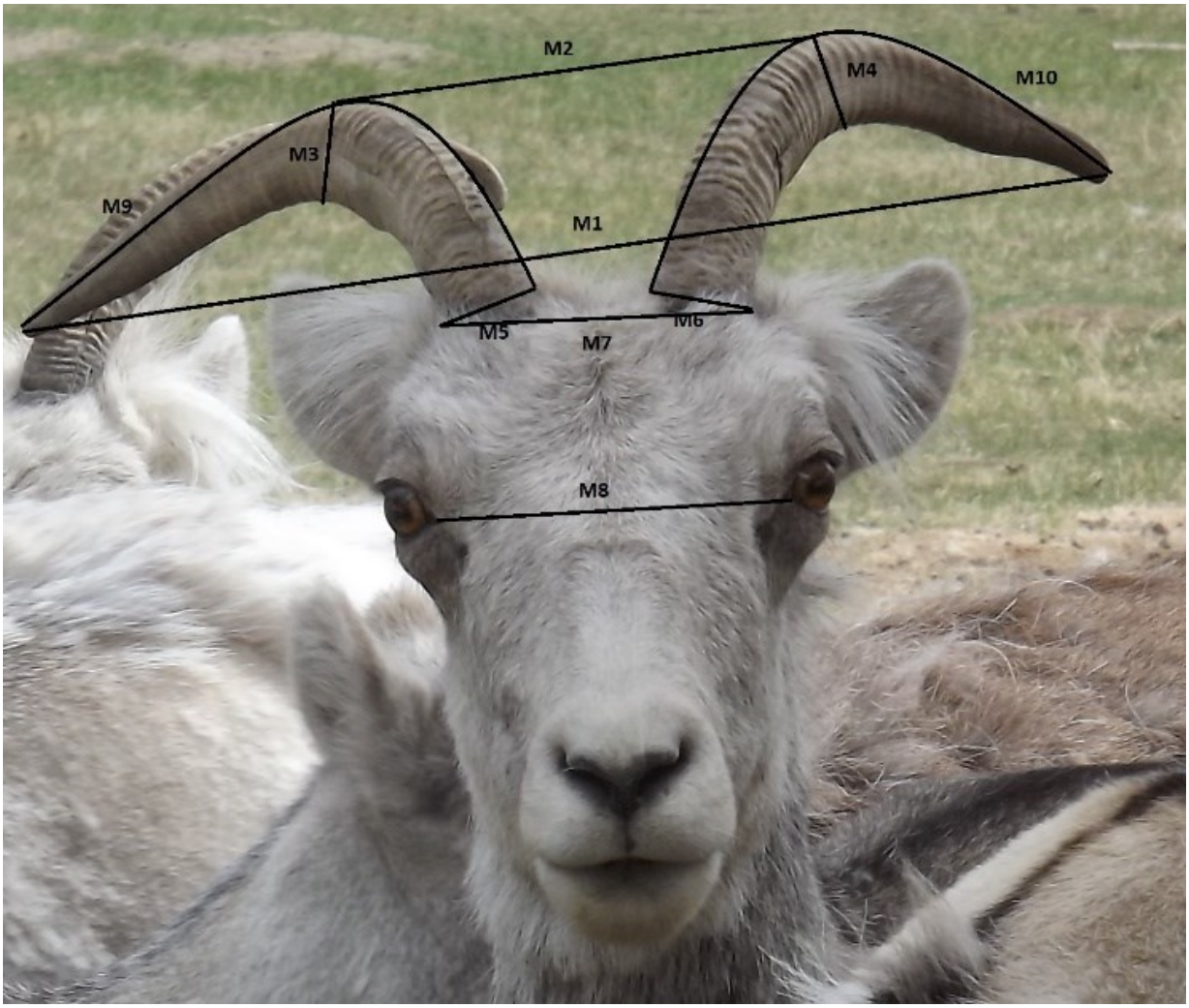

Table 7: Descriptions of 10 measurements (M1-M10) taken in the program ImageJ and used for sheep analysis 1 & 2. Measurements were modelled on those used by Merkle & Fortin (2013) for bison.

M1	Right horn tip to left horn tip
M2	Apex of right horn to apex of left horn
M3	Diameter of right horn at apex
M4	Diameter of left horn at apex
M5	Diameter of right horn at base
M6	Diameter of left horn at base
M7	Outside of horn bases
M8	Inner right eye to inner left eye
M9	Length of outside of right horn
M10	Length of outside of left horn

Table 8: Standard deviations and weights used to calculate similarity scores for sheep analysis 1 & 2 in MatchImage. Standard deviations (photograph error) were calculated by measuring 4-5 photographs each of 5 different animals and then a mean standard deviation (SD) of each of the ratios was calculated. Weights (measurer error) were calculated by measuring the same photograph 5 times and repeating this for 10 individuals. The mean SD was then taken for each ratio and used to calculate the weights. The weights are multiplicative such that the smallest mean SD was weighted as 1 with the rest less than 1, based on their respective sizes.

Ratio	Measurement	SD	Weight
r1	M1/M8	0.035	0.964
r2	M2/M8	0.073	0.584
r3	M3/M8	0.013	0.363
r4	M4/M8	0.011	0.444
r5	M5/M8	0.013	0.462
r6	M6/M8	0.019	0.568
r7	M7/M8	0.016	0.803
r8	(M1-M2)/M2	0.045	0.343
r9	M9/M8	0.116	1.000
r10	M10/M8	0.117	0.895

Figure 4: Twenty measurements (M1-M20) used for the sheep analysis 3. The 20 measurements were modelled on those used for deer. See Table 9 for descriptions.

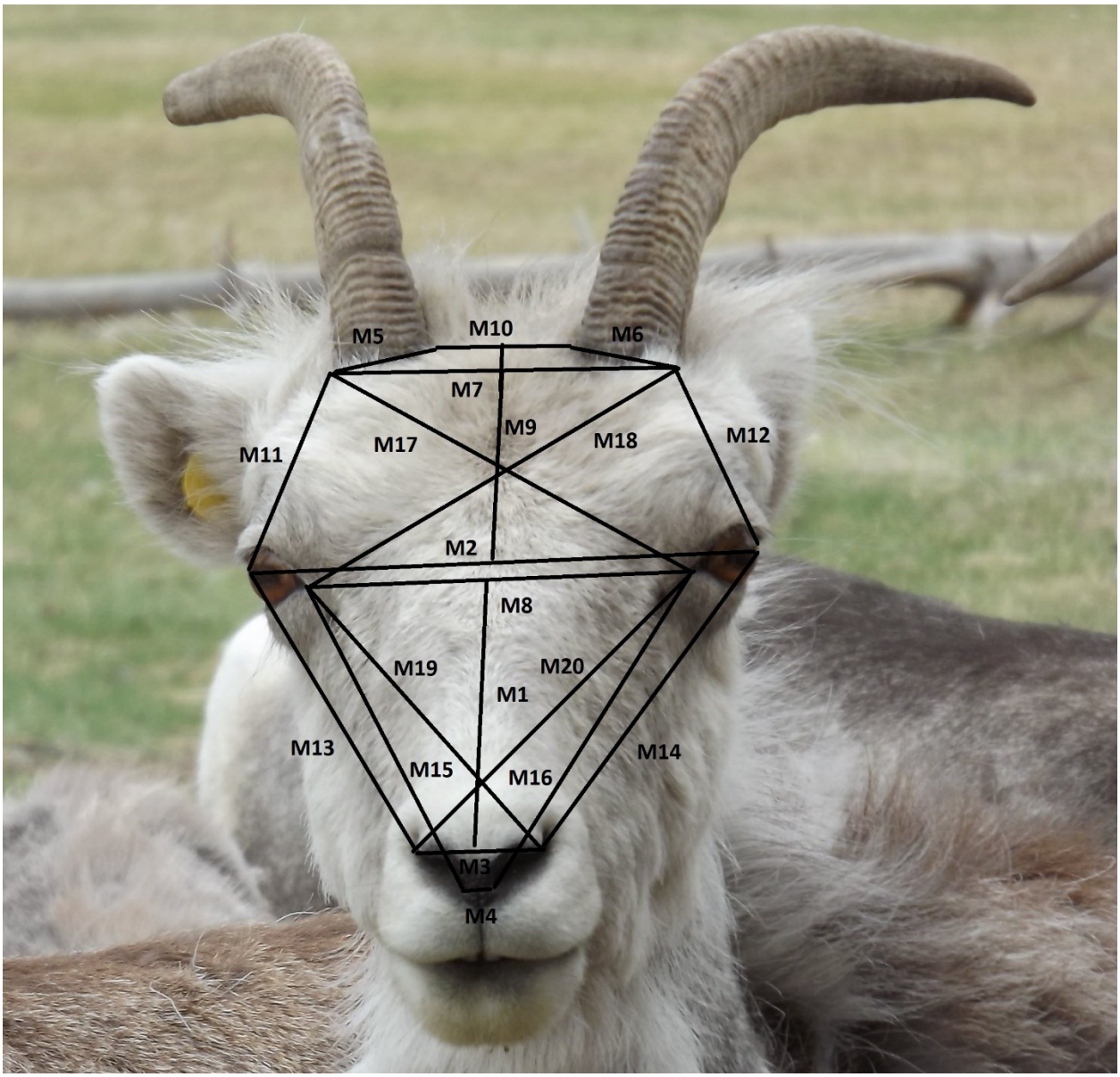

Table 9: Descriptions of 20 measurements (M1-M20) taken in the program ImageJ and used for sheep analysis 3. Measurements were modelled on those used for deer.

M1	Midpoint of M8 to midpoint of M3
M2	Outermost point of right eye to outermost point of left eye
M3	Outside of right nostril to outside of left nostril at widest point
M4	Width of septum at narrowest point
M5	Width of base of right horn at narrowest point
M6	Width of base of left horn at narrowest point
M7	Outside of horn bases
M8	Inner corner of right eye to inner corner of left eye
M9	Midpoint of M10 to midpoint of M2
M10	Inside of horn bases
M11	Left end of M5 to left end of M2
M12	Right end of M5 to right end of M2
M13	Left end of M2 to left end of M3
M14	Right end of M2 to right end of M3
M15	Left end of M8 to left of M4
M16	Right end of M8 to right end of M4
M17	Left end of M7 to right of M8
M18	Right end of M7 to left end of M8
M19	Left end of M8 to right end of M3
M20	Right end of M8 to left end M3

Table 10: Standard deviations and weights used to calculate similarity scores for sheep analysis 3 in MatchImage. Standard deviations (photograph error) were calculated by measuring 4-5 photographs each of 5 different animals and then a mean standard deviation (SD) of each of the ratios was calculated. Weights (measurer error) were calculated by measuring the same photograph 5 times and repeating this for 10 individuals. The mean SD was then taken for each ratio and used to calculate the weights. The weights are multiplicative such that the smallest mean SD was weighted as 1 with the rest less than 1, based on their respective sizes.

Ratio	Measurement	SD	Weight
r1	(M10-M7)/M7	0.131	0.157
r2	M2/M1	0.023	0.987
r3	M3/M1	0.020	0.372
r4	M4/M1	0.010	0.119
r5	M5/M1	0.022	0.432
r6	M6/M1	0.013	0.353
r7	M7/M1	0.014	0.302
r8	M8/M1	0.082	0.762
r9	M9/M1	0.032	0.509
r10	M10/M1	0.026	0.754
r11	M11/M1	0.031	0.444
r12	M12/M1	0.041	0.411
r13	M13/M1	0.105	0.821
r14	M14/M1	0.106	0.695
r15	M15/M1	0.111	1.000
r16	M16/M1	0.105	0.898
r17	M17/M1	0.057	0.886
r18	M18/M1	0.066	0.729
r19	M19/M1	0.104	0.547
r20	M20/M1	0.107	0.674

Figure 5: Ten measurements (M1-M10) used for goat analysis 1 (on the left). Measurements were modelled on those used by Merkle & Fortin (2013) for bison. Twenty measurements (M1-M20) used for goat analysis 2 (on the right). Measurements were chosen based on those used for deer. See Table 11 & 12 for descriptions.

Table 11: Descriptions of 10 measurements (M1-M10) taken in the program ImageJ and used for goat analysis 1. Measurements were modelled on those used by Merkle & Fortin (2013) for bison.

M1	Right horn tip to left horn tip
M2	Inside of horn bases
M3	Right horn base width
M4	Left horn base width
M5	Inner corner of right eye to inner corner of left eye
M6	Outermost point of right nostril to outermost point of left nostril
M7	Midpoint of M5 to midpoint of M6
M8	Outside of horn bases
M9	Length of right outer horn
M10	Length of left outer horn

Table 12: Descriptions of 20 measurements (M1-M20) taken in the program ImageJ and used for goat analysis 2. Measurements were modelled on those used for deer.

M1	Outermost point of right eye to outermost point of left eye
M2	Inside of horn bases
M3	Right horn base width
M4	Left horn base width
M5	Inner corner of right eye to inner corner of left eye
M6	Outermost point of right nostril to outermost point of left nostril
M7	Midpoint of M5 to midpoint of M6
M8	Outside of horn bases
M9	Length of right outer horn
M10	Length of left outer horn
M11	Midpoint of M9 to midpoint of M2
M12	Width of septum at narrowest point
M13	Right end of M1 to right end of M6
M14	Left end of M1 to left end of M6
M15	Right end of M5 to right end M12
M16	Left end of M5 to left end of M12
M17	Right end of M8 to right end of M1
M18	Left end of M8 to left end of M1
M19	Right end of M8 to left end of M5
M20	Left end M8 to right end of M5

Table 13: Standard deviations and weights used to calculate similarity scores for goat analysis 1 in MatchImage. Standard deviations (photograph error) were calculated by measuring 4-5 photographs each of 5 different animals and then a mean standard deviation (SD) of each of the ratios was calculated. Weights (measurer error) were calculated by measuring the same photograph 5 times and repeating this for 10 individuals. The mean SD was then taken for each ratio and used to calculate the weights. The weights are multiplicative such that the smallest mean SD was weighted as 1 with the rest less than 1, based on their respective sizes.

Ratio	Measurement	SD	Weight
r1	M1/M5	0.035	0.819
r2	M2/M5	0.015	0.376
r3	M3/M5	0.019	0.418
r4	M4/M5	0.022	0.506
r5	M8-M2/M2	0.150	0.228
r6	M6/M5	0.018	0.857
r7	M7/M5	0.051	0.994
r8	M8/M5	0.026	0.702
r9	M9/M5	0.083	1.000
r10	M10/M5	0.064	0.734

Table 14: Standard deviations and weights used to calculate similarity scores for goat analysis 2 in MatchImage. Standard deviations (photograph error) were calculated by measuring 4-5 photographs each of 5 different animals and then a mean standard deviation (SD) of each of the ratios was calculated. Weights (measurer error) were calculated by measuring the same photograph 5 times and repeating this for 10 individuals. The mean SD was then taken for each ratio and used to calculate the weights. The weights are multiplicative such that the smallest mean SD was weighted as 1 with the rest less than 1, based on their respective sizes.

Ratio	Measurement	SD	Weight
r1	M8-M2/M2	0.247	0.148
r2	M2/M5	0.030	0.996
r3	M3/M5	0.021	0.544
r4	M4/M5	0.012	0.093
r5	M5/M5	0.023	0.399
r6	M6/M5	0.019	0.446
r7	M7/M5	0.103	0.863
r8	M8/M5	0.097	0.817
r9	M9/M5	0.025	0.221
r10	M10/M5	0.102	1.000
r11	M11/M5	0.019	0.447
r12	M12/M5	0.035	0.790
r13	M13/M5	0.117	0.920
r14	M14/M5	0.108	0.911
r15	M15/M5	0.118	0.783
r16	M16/M5	0.103	0.688
r17	M17/M5	0.032	0.340
r18	M18/M5	0.030	0.287
r19	M19/M5	0.037	0.650
r20	M20/M5	0.045	0.704

Figure 6: Eighteen measurements (M1-M18) used for goat analysis 3. Measurements are analogous to those used for goat analysis 2, however, the horn length measurements M19 & M20 have been removed. See Table 15 for descriptions.

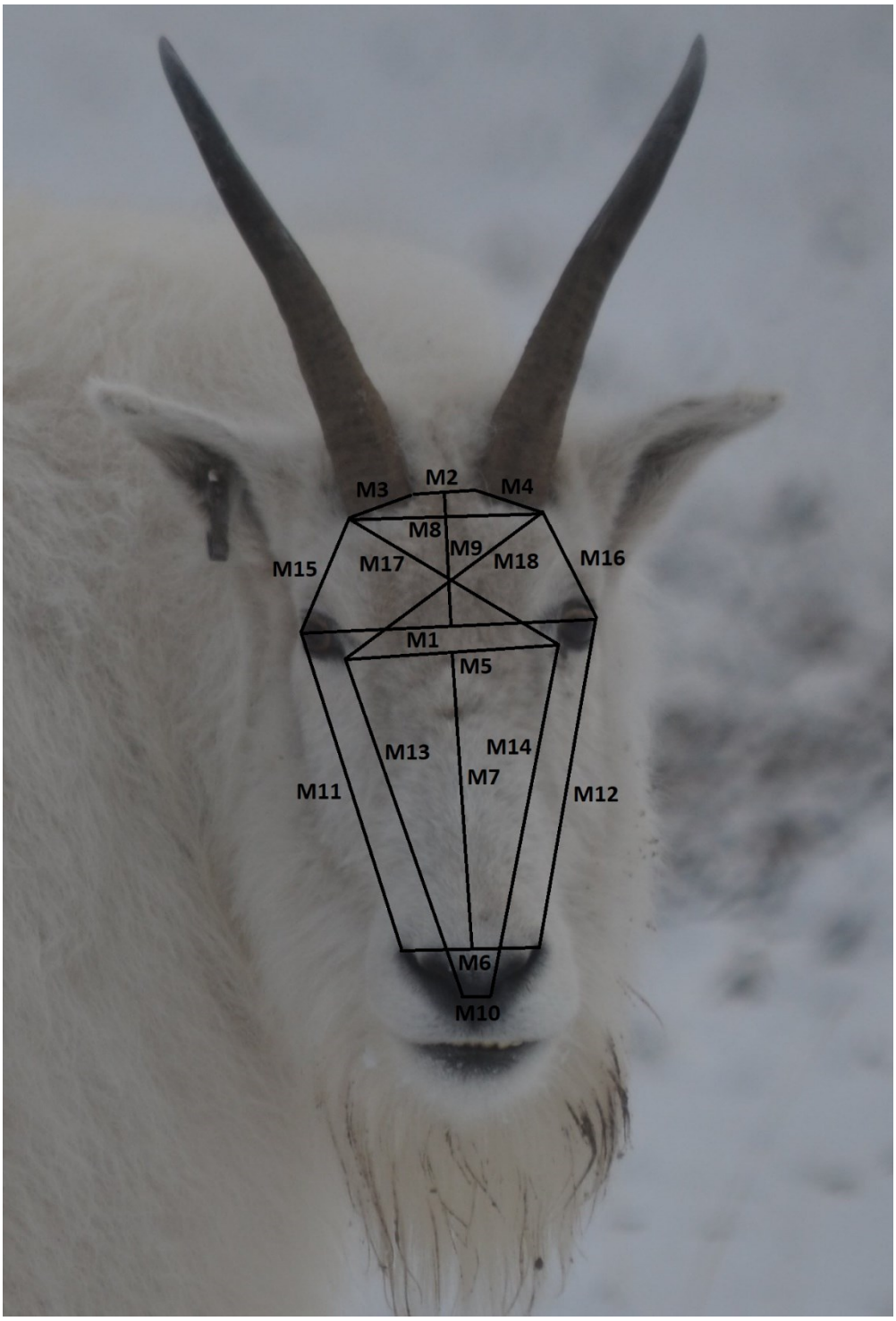

Table 15: Descriptions of 18 measurements (M1-M18) taken in the program ImageJ and used for goat analysis 3. Measurements are analogous to those used for goat analysis 2, however, the horn length measurements M19 & M20 have been removed.

M1	Outermost point of right eye to outermost point of left eye
M2	Inside of horn bases
M3	Right horn base width
M4	Left horn base width
M5	Inner corner of right eye to inner corner of left eye
M6	Outermost point of right nostril to outermost point of left nostril
M7	Midpoint of M5 to midpoint of M6
M8	Outside of horn bases
M9	Midpoint of M2 to midpoint of M1
M10	Width of septum at narrowest point
M11	Left end of M1 to left end of M6
M12	Right end of M1 to right end of M6
M13	Left end of M5 to left end M10
M14	Right end of M5 to right end of M10
M15	Left end of M8 to left end of M1
M16	Right end of M8 to right end of M1
M17	Left end of M8 to right end of M5
M18	Right end M8 to left end of M5

Table 16: Standard deviations and weights used to calculate similarity scores for goat analysis 3 in MatchImage. Standard deviations (photograph error) were calculated by measuring 4-5 photographs each of 5 different animals and then a mean standard deviation (SD) of each of the ratios was calculated. Weights (measurer error) were calculated by measuring the same photograph 5 times and repeating this for 10 individuals. The mean SD was then taken for each ratio and used to calculate the weights. The weights are multiplicative such that the smallest mean SD was weighted as 1 with the rest less than 1, based on their respective sizes.

Ratio	Measurement	SD	Weight
r1	M8-M2/M2	0.247	0.148
r2	M2/M5	0.030	0.996
r3	M3/M5	0.021	0.544
r4	M4/M5	0.012	0.093
r5	M5/M5	0.023	0.399
r6	M6/M5	0.019	0.446
r7	M7/M5	0.103	0.863
r8	M8/M5	0.097	0.817
r9	M9/M5	0.025	0.221
r10	M10/M5	0.102	1.000
r11	M11/M5	0.019	0.447
r12	M12/M5	0.035	0.790
r13	M13/M5	0.117	0.920
r14	M14/M5	0.108	0.911
r15	M15/M5	0.118	0.783
r16	M16/M5	0.103	0.688
r17	M17/M5	0.032	0.340
r18	M18/M5	0.030	0.287

Figure 7: Flow chart depicting the methodological process

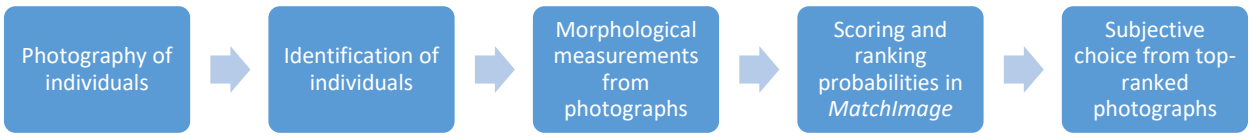

Table 17: Terms, definitions and equations

Terms	Definitions	Equation
Total false rejections (tFR)	The total number of correct matches missed by the program.	hFR+pFR
Human false rejections (hFR)	The number of correct matches that appeared in the top-ranked spots but were rejected by the user.	n/a
Program false rejections (pFR)	The number of correct matches that did not appear in the top-ranked spots.	n/a
False acceptances (FA)	The number of incorrect matches made by the user.	n/a
False rejection rate	The total number of correct matches missed by the program expressed as a rate.	(hFR+pFR)/Total number of photographs in the dataset
False acceptance rate	The total number of incorrect matches made by the user expressed as a rate.	FA/ Total number of photographs in the dataset
Average photo error	A measure of both the photo error and population variability for each measurement.	Photo error SD/Population SD
Program matching success	A measure of the accuracy of the program (including the final subjective user choice step) at achieving correct results.	{Total number of potential matches-(False rejections+False acceptances)}/Total number of potential matches
Automated matching success	A measure of the accuracy of the program had it chosen the top-ranked match in each case (without subjective user choice).	Correct matches in top-ranked spot/Total number of potential matches

Table 18: Summary of results of all analyses conducted where: tFR= total false rejections, hFR= human false rejections, pFR=program false rejections, FA=false acceptances, FRR=false rejection rate (tFR/Number of Photos), FAR=false acceptance rate (FA/Number of Photos), Actual sample size=number of individuals in the dataset, Estimated Sample Size=number of individuals identified by the program, Number of Photos=number of photographs in the dataset, Total Mis-ID=total misidentifications (tFR+FA), Number of Matches=(Number of Photos- the number of known individuals identified in the dataset), Matching Success=(Number of Matches-Total Mis-ID)/Number of Matches, Auto Matching Success=number of correct matches in the top-ranked position/Number of Matches

Analysis	tFR	hFR	pFR	FA	FRR	FAR	Actual Sample Size	Estimated Sample Size	Number of Photos	Total Mis-ID	Number of Matches	Matching Success	Auto Matching Success
Muskox 1	0	0	0	1	0.00	0.03	16	15	31	1	26	0.96	0.46
Sheep 1	8	0	8	1	0.11	0.01	32	41	76	9	44	0.80	0.52
Sheep 2	2	0	2	1	0.04	0.02	19	21	45	3	26	0.88	0.77
Sheep 3	28	2	26	6	0.25	0.05	35	62	110	34	89	0.62	0.28
Goat 1	16	1	15	4	0.24	0.06	19	37	67	20	49	0.59	0.37
Goat 2	5	1	4	3	0.11	0.06	19	26	47	8	41	0.80	0.22
Goat 3	13	2	11	4	0.28	0.09	19	31	47	17	41	0.59	0.15
Deer 1	8	1	7	13	0.13	0.22	30	36	60	21	40	0.48	0.25
Deer 2	11	7	4	6	0.22	0.12	30	35	50	17	40	0.58	0.13
Deer 3	2	0	2	0	0.11	0.00	8	10	18	2	10	0.80	0.20

Figure 8: Average photo error and FRR/FAR for all analyses. For each analysis, the population SD of each ratio was calculated and compared to the photo error SD of each ratio. To calculate average photo error, the photo error SD for each ratio was divided by the population SD to determine the percent photo error of each ratio. These were then averaged to calculate the average photo error for each analysis. The degree of photo error was categorized in Merkle and Fortin (2013) as low=10%, medium=30% and high=50%, expressed as a percent of the population SD. To calculate the FRR, the tFR was divided by the number of photographs in the database. The FAR was calculated by dividing the number of FA's by the number of photographs in the database. An increase in photo error correlates with an increase in FRR/FAR.

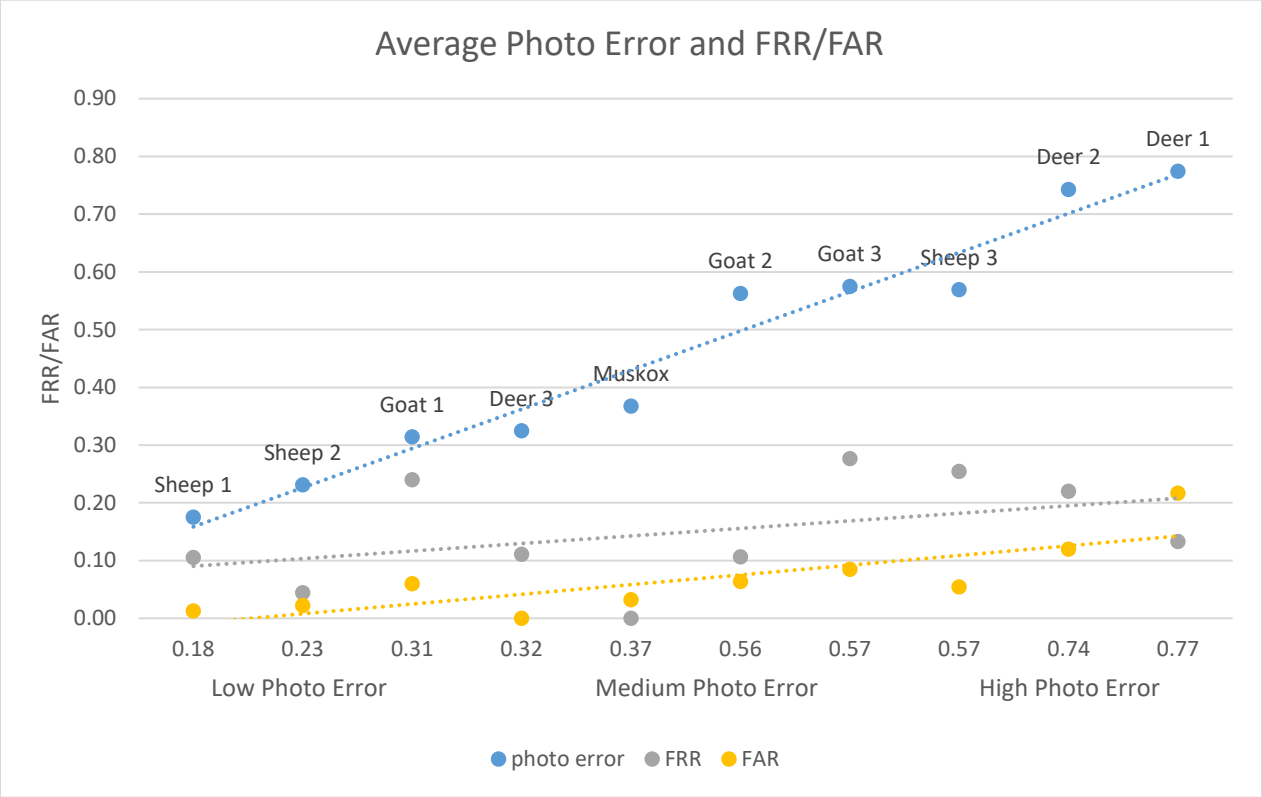

Figure 9: Comparison between automated matching success and program matching success. The automated matching success was calculated for each analysis by counting the number of correct matches in the top-ranked spot and dividing that by the number of matches (number of matches=number of photos in database minus the number of known individuals). The program matching success was determined by subtracting the tFR and the FA's from the number of matches and then dividing that by the number of matches. Program matching success rates were significantly higher than automated matching success rates (t_9=7.2, $p<0.05$).

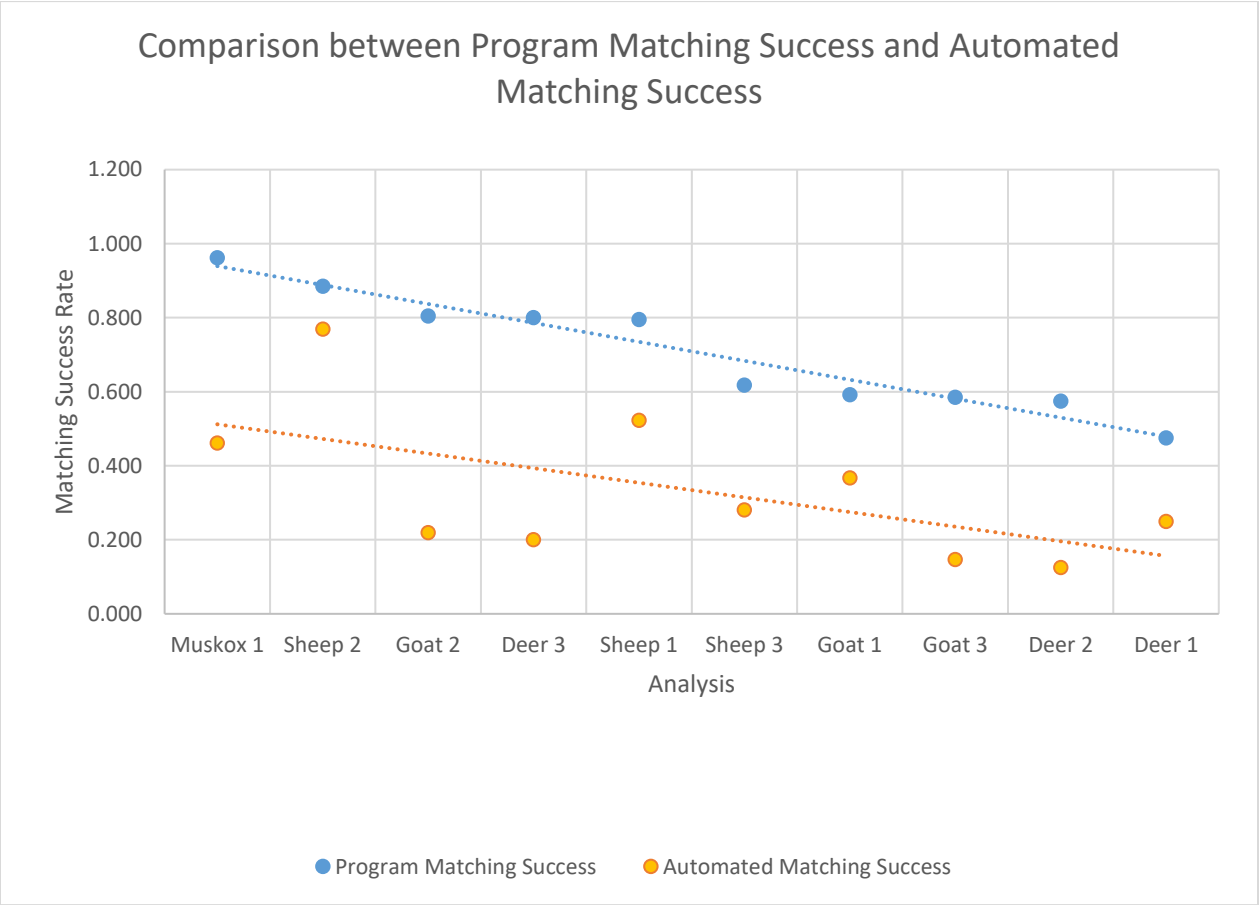

Figure 10: Examples of photographs of the same sheep taken in the fall (November) and the spring (May) from sheep analysis 1. Photo A & B: individual 1, photo C & D: individual 2, photo E & F: individual 3, photo G & H: individual 4.

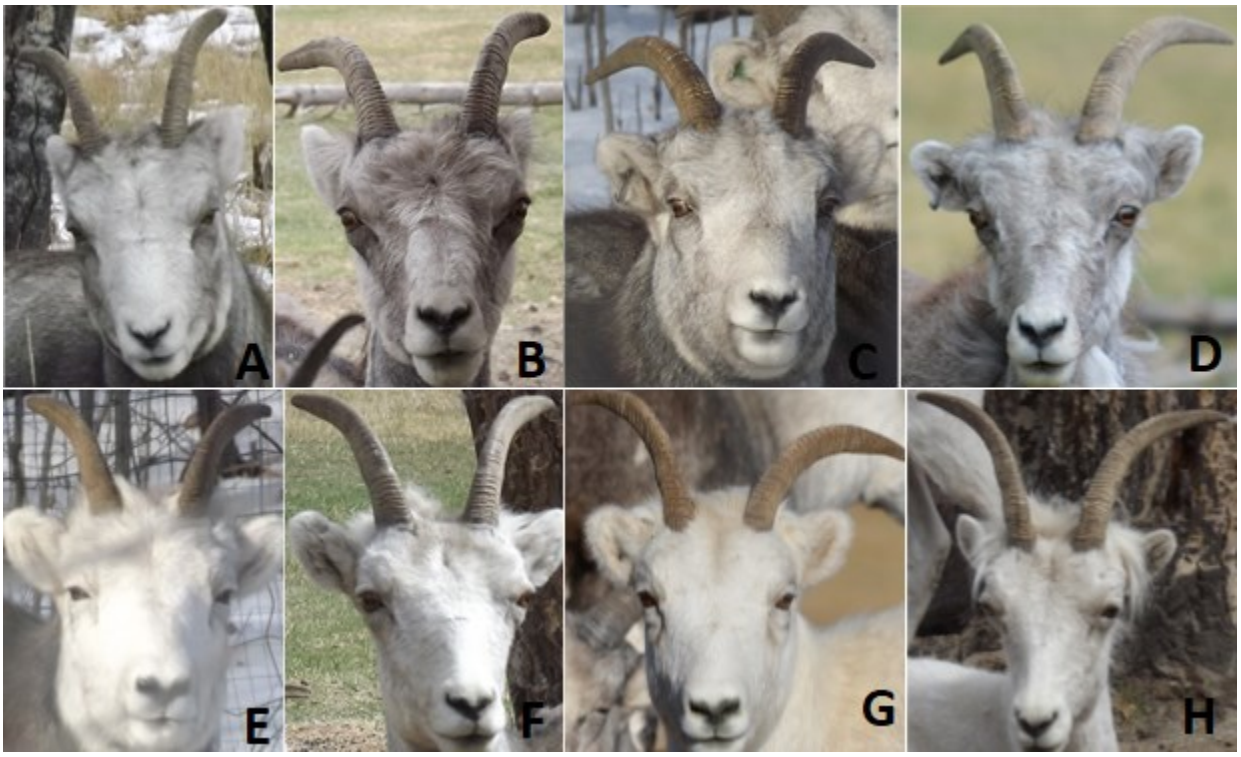

Figure 11: Deer photographs removed from the first set of photographs due to turned ears and obscured shots. These photographs were subsequently removed from the first analysis and not included in the set of photographs for second analysis.

Figure 12: Muskox photographs included in the analysis that are not head-on or at bad angles. These photographs contributed to a relatively high degree of average photo error (0.37) for the muskox analysis.

Figure 13: Goat photographs from the datasets used that were at poor angles, a) and c), too blurry to see horn rings, b), d), e) or obscured f).