J. Dairy Sci. 101:1–14
https://doi.org/10.3168/jds.2017-13768

Performance of human observers and an automatic 3-dimensional computer-vision-based locomotion scoring method to detect lameness and hoof lesions in dairy cows

Andrés Schlageter-Tello,*[1] Tom Van Hertem,† Eddie A. M. Bokkers,‡ Stefano Viazzi,† Claudia Bahr,§ and Kees Lokhorst*#
*Wageningen UR Livestock Research, PO Box 338, 6700 AH, Wageningen, the Netherlands
†Division Measure, Model and Manage Bioresponses, KU Leuven, PO Box 2456, 3001 Heverlee, Belgium
‡Animal Production Systems Group, Wageningen University, PO Box 338, 6700 AH, Wageningen, the Netherlands
§Agrifirm Innovation Center B.V., Landgoedlaan 20, 7302 HA, Apeldoorn, the Netherlands
#Van Hall Larenstein University of Applied Science, PO Box 1528, 8901 BV, Leeuwarden, the Netherlands

ABSTRACT

The objective of this study was to determine if a 3-dimensional computer vision automatic locomotion scoring (3D-ALS) method was able to outperform human observers for classifying cows as lame or nonlame and for detecting cows affected and nonaffected by specific type(s) of hoof lesion. Data collection was carried out in 2 experimental sessions (5 mo apart). In every session all cows were assessed for (1) locomotion by 2 observers (Obs1 and Obs2) and by a 3D-ALS; and (2) identification of different types of hoof lesions during hoof trimming (i.e., skin and horn lesions and combinations of skin/horn lesions and skin/hyperplasia). Performances of observers and 3D-ALS for classifying cows as lame or nonlame and for detecting cows affected or nonaffected by types of lesion were estimated using the percentage of agreement (PA), kappa coefficient (κ), sensitivity (SEN), and specificity (SPE). Observers and 3D-ALS showed similar SEN_{lame} values for classifying lame cows as lame (SEN_{lame} comparison Obs1-Obs2 = 74.2%; comparison observers-3D-ALS = 73.9–71.8%). Specificity values for classifying nonlame cows as nonlame were lower for 3D-ALS when compared with observers ($SPE_{nonlame}$ comparison Obs1-Obs2 = 88.5%; comparison observers-3D-ALS = 65.3–67.8%). Accordingly, overall performance of 3D-ALS for classifying cows as lame and nonlame was lower than observers (Obs1-Obs2 comparison $PA_{lame/nonlame}$ = 84.2% and $\kappa_{lame/nonlame}$ = 0.63; observers-3D-ALS comparisons $PA_{lame/nonlame}$ = 67.7–69.2% and $\kappa_{lame/nonlame}$ = 0.33–0.36). Similarly, observers and 3D-ALS had comparable and moderate SEN_{lesion} values for detecting horn (SEN_{lesion} Obs1 = 68.6%; Obs2 = 71.4%; 3D-ALS = 75.0%) and combinations of skin/horn lesions (SEN_{lesion} Obs1 = 51.1%; Obs2 = 64.5%; 3D-ALS = 53.3%). The $SPE_{nonlesion}$ values for detecting cows without lesions when classified as nonlame were lower for 3D-ALS than for observers ($SPE_{nonlesion}$ Obs1 = 83.9%; Obs2 = 80.2%; 3D-ALS = 60.2%). This was translated into a poor overall performance of 3D-ALS for detecting cows affected and nonaffected by horn lesions ($PA_{lesion/nonlesion}$ Obs1 = 80.6%; Obs2 = 78.3%; 3D-ALS = 63.5% and $\kappa_{lesion/nonlesion}$ Obs1 = 0.48; Obs2 = 0.44; 3D-ALS = 0.25) and skin/horn lesions ($PA_{lesion/nonlesion}$ Obs1 = 75.1%; Obs2 = 75.9%; 3D-ALS = 58.6% and $\kappa_{lesion/nonlesion}$ Obs1 = 0.35; Obs2 = 0.42; 3D-ALS = 0.10), when compared with observers. Performance of observers and 3D-ALS for detecting skin lesions was poor (SEN_{lesion} for Obs1, Obs2, and 3D-ALS <40%). Comparable SEN_{lame} and SEN_{lesion} values for observers and 3D-ALS are explained by an overestimation of lameness by 3D-ALS when compared with observers. Thus, comparable SEN_{lame} and SEN_{lesion} were reached at the expense high number of false positives and low $SPE_{nonlame}$ and $SPE_{nonlesion}$. Considering that observers and 3D-ALS showed similar performance for classifying cows as lame and for detecting horn and combinations of skin/horn lesions, the 3D-ALS could be a useful tool for supporting dairy farmers in their hoof health management.
Key words: automatic detection, cattle, hoof lesion, lameness, locomotion score

INTRODUCTION

Lameness is considered a major welfare problem in modern dairy farms. Lameness is highly prevalent with an average prevalence of 37% in England and Wales (Barker et al., 2010), 33% in Austria and Germany (Dippel et al., 2009), and from 21 to 55% in the United States (Cook, 2003; Espejo et al., 2006; von Keyser-

Received August 30, 2017.
Accepted February 24, 2018.
[1]Corresponding author: andres_schlageter@yahoo.com

lingk et al., 2012). Lameness has been associated with a reduced 305-d milk production (Warnick et al., 2001; Archer et al., 2010), a higher SCC (Archer et al., 2011), a decreased expression of estrus behavior (Walker et al., 2008), and a prolonged lapse between calving to first service and between first service and conception (Barkema et al., 1994).

Lameness is defined as impaired locomotion. The most used methods for lameness assessment in dairy cattle are manual locomotion scorings, which are procedures used to evaluate the quality of the locomotion of cows (Whay, 2002; Flower and Weary, 2009; Schlageter-Tello et al., 2014b). When scoring locomotion, observers focus their attention on gait and posture traits that are described in the protocol of the applied locomotion scoring method. Using these traits, observers assign a locomotion score to cows according to a pre-determined scale.

Hoof health management planning, in which locomotion scoring plays a crucial role, involves several steps. First, each cow is observed to evaluate gait and posture traits to assign a score for the quality of locomotion. This is usually done on a multilevel ordinal scale running from normal to severely impaired locomotion. Second, cows are classified as lame or nonlame when a predetermined threshold on the scale is exceeded, usually the middle level of the scale. It is commonly assumed that cows classified as lame suffer pain due to either hoof or other limb lesions (Flower and Weary, 2009; Schlageter-Tello et al., 2014b). Therefore, manual locomotion scoring methods are also used to detect hoof or other limb lesions (step 3). In this regard, manual locomotion scoring systems have been included in programs aimed at improving hoof health (DairyCo., 2007; Alberta Dairy Hoof Health Project, 2014) and animal welfare assessment protocols (University of Bristol, 2004; Welfare Quality, 2009). The final step within lameness management involves the choice between an appropriate treatment strategy or culling.

When using manual locomotion scoring methods to identify lameness, it is important that the locomotion scores assigned are reliable and consistent within and between observers under different practical conditions to create accurate and comparable records. In addition, if lameness is used as a visual sign for hoof lesions, it is important that cows classified as lame are indeed affected by hoof lesions. Recently, some studies questioned both the capability of human observers to perform locomotion scoring consistently and the utility of lameness for lesion detection (Engel et al., 2003; Tadich et al., 2010; Schlageter-Tello et al., 2014b).

In recent years, several automatic locomotion scoring systems have been developed due to the increasing number of animals per dairy farm and to the lack of time on the part of the farmers to monitor the increasing number of animals or to improve methods for better detection of lameness and hoof lesion (Rutten et al., 2013; Schlageter-Tello et al., 2014b; Van Nuffel et al., 2015). Most automatic locomotion scoring systems attempt to mimic human observers by measuring and analyzing parameters of cows' locomotion and behavior through sensors and mathematical algorithms. Some examples include measuring forces exerted on the floor by the limbs using force plates (Rajkondawar et al., 2002) or 3-dimensional (**3D**) force plates (Dunthorn et al., 2015), weight distribution of limbs using 4 independent weighing units (Chapinal et al., 2009a), parameters associated with distances between hoof prints using pressure-sensitive mats (Maertens et al., 2011), or parameters associated with activity and behavior using accelerometers attached to the neck or limbs of cows (Alsaaod et al., 2012; Thorup et al., 2015). Recently a promising approach for an automatic locomotion scoring used 3D camera technology to measure different angles associated with back curvature (Viazzi et al., 2013; Van Hertem et al., 2014). The advantages of the 3D computer vision automatic locomotion scoring system (**3D-ALS**) include utilization of a single sensor (1 camera) to assess locomotion in a large number of cows, the same set-up may be used to assess different parameters (e.g., BCS) and it shows acceptable performance for lameness detection (Viazzi et al., 2013; Van Hertem et al., 2014).

The 3D-ALS and most automatic locomotion scoring methods are evaluated for lameness detection using a locomotion score or lame/nonlame classification assigned from observers to a cow as a reference (Schlageter-Tello et al., 2014b). Most studies, however, comparing automatic and manual locomotion scoring report only the performance of the automatic systems compared with observers performing locomotion scoring, but do not report the performance of observers used as reference (Schlageter-Tello et al., 2014b). Similarly, few studies compare the performance of manual and automatic locomotion scoring systems with presence/absence of hoof lesions, and to our knowledge, only one article compared the performance of both manual and automatic locomotion scoring systems for detecting hoof lesions under the same practical conditions (Bicalho et al., 2007). Thus, an actual comparison between both systems for lameness assessment and hoof lesion detection has not yet been performed properly.

Given the lack of information when comparing both manual and automatic locomotion scoring, the objective of this study was to determine if a 3D-ALS was able to outperform human observers performing manual locomotion scoring for classifying cows as lame/nonlame and detecting specific types of hoof lesions.