The 2009 IEEE/RSJ International Conference on
Intelligent Robots and Systems
October 11-15, 2009 St. Louis, USA

Human Robot Interaction
Studies on Laban Human Movement Analysis and Dynamic Background Segmentation

Luís Santos, José Augusto Prado, Jorge Dias

Abstract— **Human movement analysis through vision sensing systems is an important subject regarding Human-Robot interaction. This is a growing area of research, with wide range of aplications fields. The ability to recognize human actions using passive sensing modalities, is a decisive factor for machine interaction. In mobile platforms, image processing is regarded as a problem, due to constant changes. We propose an approach, based on *Horopter* technique, to extract Regions Of Interest (ROI) delimiting human contours. This fact will allow tracking algorithms to provide faster and accurate responses to human feature extraction. The key features are head and both hand positions, that will be tracked within image context. Posterior to feature acquisition, they will be contextualized within a technique, *Laban Movement Analysis* (LMA) and will be used to provide sets of classifiers. The implementation of the LMA techquine will be based on Bayesian Networks. We will use these bayesian classifiers to label/classify human emotion within the context of expressive movements. Compared to full image tracking, results improved with the implemented approach, the *horopter* and consequently so did classification results.**

I. INTRODUCTION

Human movement analysis is an active area of research with a growing number of surveys. Gavrila [4] points out that the ability to recognize human movements and their activities using passive sensing modalities, namely vision, is a key factor for machine interaction in an inteligently and effortlessly fashion within a human-inhabited environment. Human-machine interaction can beneficiate from this simple sensor, allowing 'communication' to become easier and uncumbersome. This work uses a technique, *Laban Movement Analysis* (LMA), to analyze human gestures. A classifier was implemented based on LMA, which uses low-level-features (LLF) as the very basic level to characterize human emotion within the context of expressive movements. These descriptors will be extracted from spatial trajectories generated by specific body parts along the time. The key body parts to be tracked are the head and both hands. To acquire the body part trajectories, two sensing modalities are used: vision and magnetic sensor. However, in mobile human-interacting platforms, image processing can pose as problem, due to dynamic characteristics of the elements in the image. This work proposes the use of the *geometric horopter* defined by the vision system to extract, via signal processing, Regions Of Interest (ROI) delimiting human contours. As a consequence, the tracker will no longer search the entire image, rather it will be confined to the computed ROI. This fact will improve the tracking algorithm speed and accuracy to body part position extraction.

This work is partially supported by the BACS-project-6th Framework Programme of the European Commission contract number: FP6-IST-027140.

L. Santos and J. Augusto and J. Dias are with the Institute of Systems and Robotics, Department of Electrical Engeneering, Pólo II, 3000 Coimbra, Portugal (luis, jaugusto, jorge)@isr.uc.pt

II. BASIC CONCEPTS

A. Laban Notation and Bayesian Approach

Laban Movement Analysis (LMA) is a method to observe, describe, notate and interpret human movement, developed by Rudolf Laban (1879 to 1958). The general framework is widely applied in physical and menthal therapy [1] as well as studies of dance, however, it has found little application in the engineering domain. A recent study by Rett J. [8], explored how LMA can be used to classify human expressive movements within human-machine interaction. A robot interface was developed with the capability to interpret a set of basic human movements. Rett's work also states that LMA has the potential to analyze emotional content of human actions.

Laban theory consists of several major components, though the available literature does not set a standard regarding their total numbers. The work of Norman Badler's group [10, 2] mentions five major components.

Non kinematic components: *Body* specifies which body parts are moving, their relation to the body center; *Space* deals directly with the trajectory executed by the body parts while performing a movement. Within the **Kinematic** ones there are: *Effort* which deals with the dynamic qualities of the movement, and the inner attitude towards the use of energy; *Shape* (emerging from *Body* and *Space*) is focused on the body itself. The fifth component is *Relantionship* that appears as the less explored one, and describes the interaction with oneself, others and the environment. Some literature only considers the first four mentioned components [3]. The *Space* component has already been studied by Rett et. al. [6], though in his recent study [8], space was left for improvement on the *Effort* component, in order to study the emotional content of expressive movements. The *Effort* component will be the main focus within movement analysis in this work, hence a short description will follow.

Effort

What makes the framework of LMA so special is its ability to describe an additional 'expression' that accompanies the spatial trajectory (*Space* component). By retrieving some

TABLE I

Effort QUALITIES AND THEIR SUBJECTS

Effort	Cognitive process	Subject	Extremes
Space	Attention	The spatial orientation	focused or non-focused
Weight	Intention	The impact	strong or light
Time	Decision	The urgency	urgent or non-urgent
Flow	Progression	How to keep going	free or careful

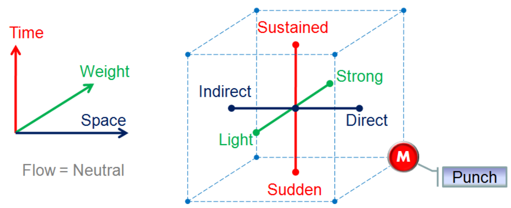

Fig. 1. The bipolar *Effort* qualities of the *Action Drive*, i.e. *Flow* = *neutral* (omitted) represented as a cube. The position of the movement M (*punch*) indicates its qualities, i.e. *direct*, *sudden* and *strong*.

evidences about the emotional state or the intention of the performer, the *Effort* component can be seen as the key descriptor to solve the task of analyzing 'expressive movements'.

Effort describes the dynamic qualities of the movement and the inner attitude towards using energy. It consists of four *Effort* qualities: *Space*, *Weight*, *Time*, and *Flow*. Table I shows the *Effort* qualities, the underlying cognitive process, the subject and the two extremes each quality has [1]. Each quality is bipolar and lies between two extremes. The values for the *Effort* qualities are shown in (1)

$$\begin{aligned} Space &\in \{direct, neutral, indirect\} \\ Time &\in \{sudden, neutral, sustained\} \\ Weight &\in \{strong, neutral, light\} \\ Flow &\in \{bound, neutral, free\} \end{aligned} \tag{1}$$

Movements are described and distinguished by those qualities close to an extreme, e.g. a *Punch* has *Strong Weight*, *Sudden Time* and *Direct Space*.

Combinations of all four qualities rarely occur as they produce extreme movements (e.g. tearing something apart) [1]. Single-quality movements are also rare [1] [10]. Combinations of three qualities, with the fourth considered to be neutral, appear to be the most natural way to perform an action. These combinations of three *Effort* qualities can be divided into four cathegories: *Action Drive, Weightless, Timeless* and *Spaceless*, considering as being neutral, the *Flow, Weight, Time* and *Space* respectively. Each of these combinations, can be modulated to define a 3-D space, a cube where each vertex represents a sub-component (Fig. 1).

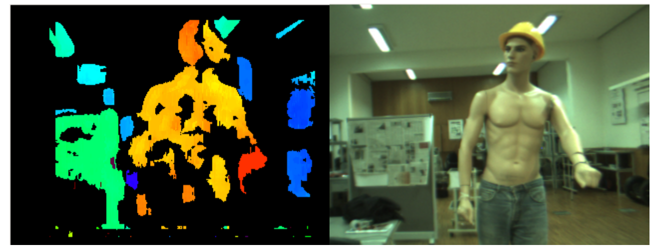

Fig. 2. a) Depth map (*'hot'* colors represent nearest areas, *'cold'* colors represent further ones; b) Dominant eye raw image

B. Interaction Zone and Dynamic Backgrounds

In order to human-robot interaction occur, the robot needs to identify the subject with whom it will interact. This process usually involves video sensing.

The system is implemented in a mobile platform, and the use of a monocular vision system has inherent problems. The extraction of a person from the image demands more complex image processing due to dynamic characteristics (other people and/or skin color like objects) of the background. Algorithms based on haar-like features and color filters have their application compromised. This means that an approach based on static background, as in [9] and [7], is not possible.

The challenge was thus to have a robust real time solution for dynamic background segmentation on mobile robotics. It was decided to base our approach on the *Geometric Horopter*.

The first step of this technique uses stereo vision to generate a *depth map*. In Fig.3 a) the *depth* map resulting from the application of this algorithm is presented, whilst the right side shows the image from one of the stereo cameras.

Within the application of the *horopter* a new definition is introduced, the *interaction zone*. A circumference will be defined from the geometric setup of the system, and its inner area defines the interaction zone. Any subject that intents to interact with the system must lie inside this circumference. The geometry behind this definition is explained in the following section.

C. Geometrical Horopter

1) Horopter Segmentation :

a) Properties of ViethMuller Circle: The concept of *interaction zone* has been defined as dependent of a circle. That circle is called the Vieth-Muller Circle, where the following properties can be defined (See Fig.4):

- In a pure version eye movement, the fixation point moves along the same ViethMuller Circle. Fig.4 a) illustrates this fact showing how P moves to P' along the Vieth-Muller Circle.
- If the fixation point remains static, the disparity for various points is studied. Disparity is defined as ϕL ϕR.

The Vieth-Muller circle is geometrically defined from two points and a distance. The two points are easily found, as

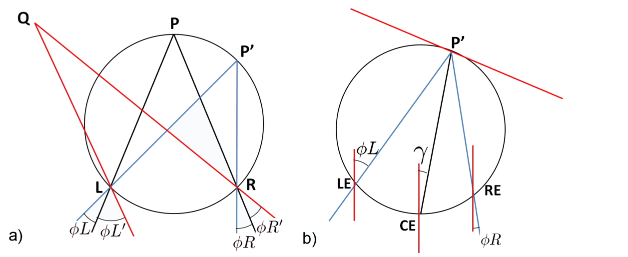

Fig. 3. a) Calculating the Disparity; b) Disparity Properties on Vieth-Muller Circle

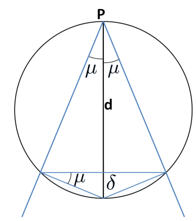

Fig. 4. Simples justification scheme for value γ

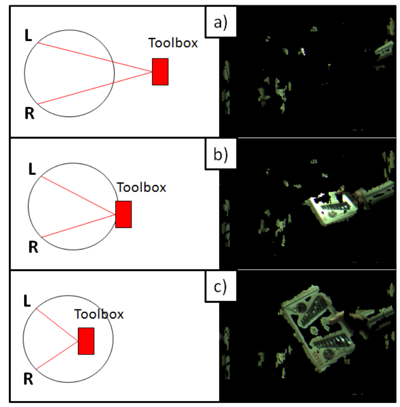

Fig. 5. a) The toolbox is yet outside the Vieth-Muller Circle; b) Toolbox starting to enter the horopter zone; c) The object is fully inside the Vieth-Muller circle, and thus, visible.

they reflect the camera positions. The distance will define the interaction zone, making the selection criteria, an important aspect. Several options for defining this distance could be chosen, however a simple approach was selected: to use a fix value of 2 meters.

b) Theorem 1: If a point Q lies on ViethMuller Circle, its disparity is zero.
As Q moves outside (e.g. point P moves to position Q in Fig.4 a)), ϕL decreases whilst ϕR will naturally increase. However ff point Q moves inside the circle, the opposite relation between ϕL and ϕR occurs.

c) Theorem 2: Disparity is nonzero outside the circunference line of the Vieth-Muller Circle (with opposite signals, depending on whether side of the circle it lies in, outside or inside).
For human vision system, when the disparity has high enough values, the object is seen in double (one from left eye and the other from right eye). This phenomenon is called *Diplopia*. The maximum disparity prior to the diplopia even is defined as *Panum's Fusional Limit*.

d) Calculating Disparity: The ϕL and ϕR are made by line of sight with the straight ahead direction. The $GazeAngle$ γ (see Fig.4 a) and $VergenceAngle$ μ (see fig. 4) are defined as

$$\begin{aligned} \gamma &= \frac{1}{2}(\phi L + \phi R) \\ \mu &= \frac{1}{2}(\phi L - \phi R) \end{aligned} \tag{2}$$

CE represents the cyclopean eye and $(d + \delta)$ is the distance from CE to the target object (see fig. 4).

The Horizontal Disparity is

$$h = \frac{I \cos\gamma}{d}\left(\frac{\delta}{\delta + d} + \frac{d \tan\gamma}{\delta + d}x + x^2\right)$$

and Vertical Disparity

$$v = \frac{I \cos\gamma}{d}\left(\frac{d \tan\gamma}{\delta + d}y + xy\right)$$

where (x, y) are cyclopean image coordinates and I is the interocular distance.

e) Theorem 3: $d = I \cos\gamma / \sin 2\mu$
A simple justification can be presented for the value of $\gamma = 0$, as it can bee seen in Fig.4.

$$I/2 = d \times \sin u \times \cos u \Rightarrow d = I \cos r / \sin 2u$$

The disparity is calculated, and the resulting depth image (Fig.3 a)) is correlated with the CE image.

The pixel values in the final CE image will depend on the disparity values: pixels with negative disparity values will be assigned the value '0'; for positive disparity values, the pixel maintains its original value. The result is a masked image, where only 'visible' objects within the circle (the interaction zone) can be seen (right column of figure 5).

The resulting image allows us to define a Region Of Interest (ROI) that can be approximated for an elipsoid. This image will be further processed for *face/hand* detection. The authors present this approach as a robust way to ensure that the robot will interact only with visible subjects in the image, i.e. subject that are inside *Vieth-Muller*circle.

III. Human Robot Interaction

A. Database of Expressive Movements

According to LMA Effort definitions of *Action drive, Timeless, Weightless* and *Spaceless*, there are 32 different possible combinations. This work is focused on *Effort* component (work on *Space* component has been done in [6]), thus *Space* component has no particular relevance in this set of results. The database was build around movements

TABLE II

Basic Effort Weightless

Action	Example	E.Sp	E.Fl	E.Ti
Punch	Forward punch	Dir	Bound	Sud
Writting	Write name with a spray can	Ind	Free	Sus
Lift	Lift heavy object	Dir	Bound	Sus
Flick	Clean with a brush	Ind	Bound	Sud
Free scene	Free movements	misc	misc	misc

that present certain *Effort* characteristics, in a way that will allow the *Bayesian Classifier* to accurately provide an *Effort* labelling of the movements, based on LMA. Table II shows a set of *Weightless* expressive movements and their qualities.

The data set also encompasses a free scene hand labeled, were several of the *Effort* characteristics were performed along the time.

B. Tracking

Data acquisition is processed using two different types of sensor:

1) a 6-DoF magnetic tracker that provides 3-D coordinates with high accuracy and speed (50Hz),
2) a regular firewire camera.

In the learning stage, to train the bayesian framework, the data provided by the magnetic tracker is used to provide accurate data to work as "'ground truth"'. However, in real life situations, it is not feasible to have magnetic sensors attached in subjects, thus in the classification stage vision sensing was used as the data acquisition method.

The magnetic tracker has the possibility to work eight sensors at a time. However in the current work, only three were used, which are positioned in the head, and one at each hand of the performer. The tracker provides the raw position data for each of the body parts, relative to the magnetic tracker referential. To acquire the body part position using the vision system, a simple algorithm was implemented. The detailed description of the tracking algorithm is out of the scope of this work, however a brief description will be given. The algorithm starts with the acquisition of the image obtained from the application of the geometric horopter (see Section II-C). Haarlike features are applied for face detection initialization. After the face has been detected, a color histogram based algorithm, CAMshift, is used to track body parts based on skin color information. For faster initialization of the color tracker, body geometric constrains use the face position to compute the initial guess of hand positions. Both Haarlike and CAMshift algorithms used belong to the OpenCV Library.

Regardless of the acquisition system used, the acquired data is a collection of hand and head coordinates in 3-D and 2-D space for magnetic tracker and camera respectively. Using geometric transformation, homography [5], the magnetic tracker 3-D coordinates are mapped into the camera referential. The homography process was done considering the calibration between camera and magnetic tracker. The process of calibration was done using projective geometry [5].

TABLE III

INITIAL HYPOTHESES OF CORRESPONDENCES BETWEEN LMA *Effort* QUALITIES AND PHYSICAL ENTITIES

LMA Effort Qualities	Physical entities
Time.sudden	**High acceleration**, High velocity
Time.sustained	**Low acceleration**, Low velocity
Space.direct	**Small curvature**, Small angular velocity
Space.indirect	**High curvature**, High angular velocity
Flow.free	**High curvature, High angular velocity**
Flow.bound	**Low acceleration, Low velocity**
Weight.strong	Muscle tension, Medium acceleration
Weight.light	Muscle relaxed

C. Prominent physical features

The selection of relevant features is one on the great mysteries in pattern recognition. In this work, features were chosen by interpreting the parameters of Laban Movement Analysis (LMA) through physical measurable entities that could describe them best. The resulting data in form of coordinates allows the computation of physical features such as velocity, acceleration and curvature of the discretized trajectories (e.g. having two sets of coordinates, it is possible, knowing the time interval between each of them, to compute the velocity from one point to the other). Having this sort of features in mind, the initial hypotheses of correspondences between LMA parameters and physical entities are expressed as shown in Table III.

It seems most natural that high velocities and accelerations are connected to sudden/fast movements. Also when trajectories exhibit low curvatures, one must assume that they are mainly linear, or in Laban parameters, *Direct*. The behavior of these features within each of the LMA qualities was studied, and the most prominent are marked in bold in table III. As it can be seen, this work still has not found physical features that can inequivocally define *Ef.Weight* (there is no visual cue that could reflect muscle tension). However, ongoing work is still being conducted focused on visual cues that could reflect it. In Fig. 6 an example of physical features behavior in a determined *Effort* quality is presented. Velocity (Fig.6a)) and curvature (Fig.6c)) do not present a strong pattern that can distinguish both extremes of *E.Ti*, however Acceleration (Fig.6b)) presents a strong, distinguishable feature.

All features are identically processed, originating the final relations for the models presented in table III.

IV. PROBABILISTIC MODELING

A. Effort Model

The *Effort model* describes the dynamic aspects of the movement. It relates the low-level features like speed (Vel), acceleration (Acc) and curvature (K) to *Effort* qualities like *Time* ($E.Ti$), *Space* ($E.Sp$), *Weight* ($E.We$) and *Flow* ($E.Fl$). In order to not confuse the *Space* **component** from the previous section with the *Space* **quality** of the *Effort* component, all variable symbols of Effort qualities

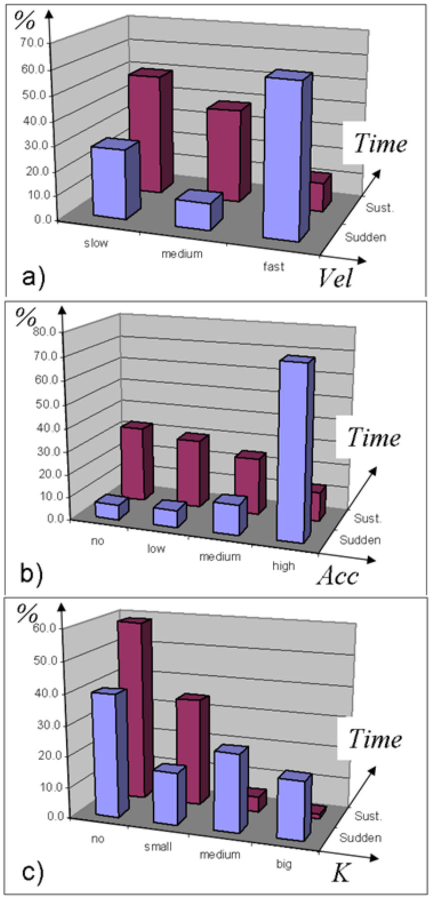

Fig. 6. Behavior of Velocity V, Acceleration A and Curvature K in *Effort Time*

are preceded by a leading E.

$$\begin{aligned} \mathbf{LLF}_{Ef} &\in \{Vel, Acc, K\} \\ \mathbf{Effort} &\in \{E.Ti, E.Sp, E.We, E.Fl\} \end{aligned} \tag{3}$$

The relation between the two sets of variables described in 3 has already been investigated, established and developed in Section III. The *concept space* relates the *Effort* qualities to a specific movement M. The *Effort model* is related with a specific plane and body part where the *Effort* qualities can be detected best. Variables and their space are shown in (4)

$$\begin{aligned} Acc &\in \{no, low, medium, high\}\ \langle 4 \rangle \\ K &\in \{no, small, medium, big\}\ \langle 4 \rangle \\ E.Sp &\in \{direct, indirect\}\ \langle 2 \rangle \\ E.Ti &\in \{sudden, sustained\}\ \langle 2 \rangle \\ E.Fl &\in \{free, bounded\}\ \langle 2 \rangle \end{aligned} \tag{4}$$

Each movement M will produce a certain set of *Effort* qualities during the movement action. Thus we have a conditional dependency of *Effort Space* $E.Sp$, *Effort Time* $E.Ti$ and *Effort Flow* $E.Fl$ from the movement M as can be seen in Bayes-net of Fig. 7.

The *Effort* variables can not be directly measured but observed through some low-level features (i.e. $\mathbf{LLF}_{Ef}$).

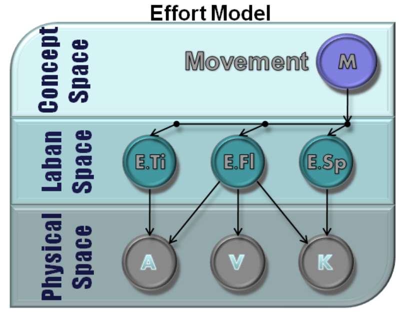

Fig. 7. Laban Effort Bayesian Model

TABLE IV
EFFORT VARIABLES

Variable	Symbol	Description
Movement	M	e.g. $Punch$
Effort Space	$E.Sp$	e.g. $E.Sp = direct$
Effort Time	$E.Ti$	e.g. $E.Ti = sudden$
Effort Flow	$E.Fl$	e.g. $E.Fl = bound$
Speed gain	Acc	e.g. $Acc = high$
Curvature	K	e.g. $K = small$
Velocity	V	e.g. $V = medium$

Thus, there is a dependency of the non-observable variables from the *Effort* set and $\mathbf{LLF}_{Ef}$. The joint distribution can be expressed as

$$\begin{aligned} & P(M\ E.Sp\ E.Ti\ E.Fl\ A\ K\ V) \\ &= P(M)\ P(E.Sp \mid M)\ P(E.Ti \mid M)\ P(E.Fl \mid M) \\ & P(A \mid E.Ti\ E.fl))\ P(K \mid E.Sp\ E.Fl)\ P(V \mid E.Fl) \end{aligned} \tag{5}$$

The variables used in this sub-model are summarized in Table IV.

V. Results and Discussion

This section will briefly describe the general setup, and will divide the results in two sections: the results of geometric horopter and consequent tracking and the results for the Effort analysis. Four movements and a free scene were processed, with 100 trials per movement, with 10 subjects (both gender), with no more than 10 trials per different person within each gesture.

A. Experimental Setup

A pair of *Firewire* cameras compose the stereo vision system. The movement sequences were acquired with Magnetic Tracker Device (Polhemus Liberty). The data acquired by the magnetic tracker was used to train our bayesian network. Classification results derived from features acquired through image video sequences which simulate real-time conditions perfectly. The analyzed movements are summed up in Sec.III-A tableII.

B. Geometric Horopter Results

The Horopter results are easily understood if visualized, hence in Fig.8 it can be seen the segmentation that occurs

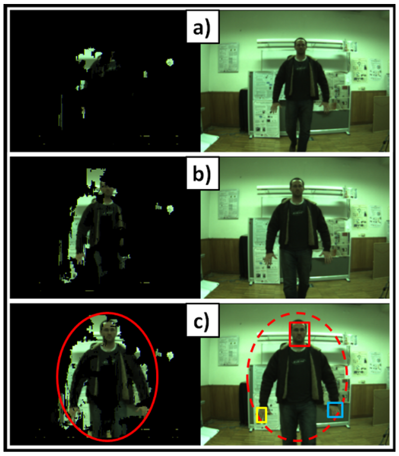

Fig. 8. a) Subject outside FoV. b) Subject entering FoV (Haarlike features try to identify a face) c) Person inside FoV, Face detected and CAMshift starts hand and head tracking

TABLE V

RESULTS FOR EFFORT QUALITIES

	Physical Qualities					
	Space		Flow		Time	
	Ind	Dir	Free	Bound	Sud	Sus
Positive results	79.3%	90.2%	61.2%	58.7%	84.8%	97.1%

with this algorithm. Once the subject enters the horopter field of view (FoV) completely, the result is a region where the person can be seen (that region tends to approximate a contour). This countor is then aproximated by an elipse (Fig.8c) left side), which defines the region where the tracking is processed (superimposed in Fig.8c) right side). Haarlike features detect a face within the delimited region (Fig.8c) right side), after which CAMshift is triggered to start acquiring information relative to hand and head positions. This algorithm is, as said, based in color histograms and kalman filtering. Cases happen where tracker is momentarily lost. In those cases, the probability distributions in each node of the bayesian network are not updated.

C. Emotion Analysis Results

The usefulness of Laban Movement Analysis within emotion classification is demonstrated with the results shown in table V. The positive results represent the percentage of trials that have achieved correct results, e.g. if in 10 trials, 7 were correcly identified and 3 were not, the result would be 70%. For *E.Time* and *E.Space* the results are considered good. However for *E.Flow* the results exhibit some confusion. This arises from the fact that low accelerations only describe the *'bound'* state of *E.Flow*, but acceleration is still counted as evidence for the *'free'* state. Similar conclusion when analyzing Curvature. *Effort* qualities, due to the number of different combinations possible, allow the possibility to distinguish movements almost based on *Effort* component. The selected features already allow a good discretization,

TABLE VI

RESULTS FOR EFFORT AND SPACE COMPONENTS COMBINED

	Laban Components		
	Space	Effort	Space+Effort
Classification Rate	61.3%	86.4%	79.4%

and when combined with the other LMA components can provide a robust tool for human movement analysis.

Previous work [6] has been done concerning the *Space* component alone, however the fusion of the *Space* and *Effort* components in one simple bayesian model defined as $P(Movement|Space.component\ Effort.component)$ showed what can be described as significant improvement. As seen in the following table, the results from [6] yielded positive results around 60%, and the fusion with the *Effort* component implemented in the current work, improved to around 80% the overall movement classification rate.

VI. CONCLUSION

From the results achieved, we can conclude that horopter is a valid approach for dynamic background segmentation, provided that it receives background with enough features, which usually happens. This segmentation enhances tracking results, both in speed and accuracy and should be further explored. Laban Movement Analysis is without a doubt a powerfull movement descriptive tool, results show that it can, with some accuracy classify basic emotion primitives (contextualized within LMA), and the implementation of the remaining components is an ongoing work. The final goal of this work, is to build an autonomous interactive social robot.

REFERENCES

[1] I. Bartenieff and D. Lewis. *Body Movement: Coping with the environment.* New York: Gordon and Breach Science., 1980. [1, 2]

[2] Diane Chi, Monica Costa, Liwei Zhao, and Norman Badler. The emote model for effort and shape. In *SIGGRAPH 00, Computer Graphics Proceedings, Annual Conference Series*, ACM SIGGRAPH, pages 173–182. ACM Press, July 2000. [1]

[3] Afra Foroud and Ian Q. Whishaw. Changes in the kinematic structure and non-kinematic features of movements during skilled reaching after stroke: A laban movement analysis in two case studies. *Journal of Neuroscience Methods*, 158:137–149, 2006. [1]

[4] D. M. Gavrila. The visual analysis of human movement: A survey. *CVIU*, 73(1):pp. 82–98, 1999. [1]

[5] R.I. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision.* Cambridge University Press, 2000. [4]

[6] J.Rett, L. Santos, and J. Dias. Laban movement analysis for multi-ocular systems. In *IEEE/RSJ International Confece on Intelligent Robots and Systems (IROS08)*, 2008. [1, 3, 6]

[7] J. Meessen, C. Parisot, C. Lebarz, D. Nicholson, and J.F. Delaigle. Smart encoding for wireless video surveillance. In *In SPIE Proc. Image and Video Communications and Processing*, volume 1, 2005. [2]

[8] Joerg Rett. *Robot-Human Interface Using Laban Movement Analysis Inside a Bayesian Framework.* PhD thesis, University of Coimbra, 2009. [1]

[9] Theodore Totozafiny, Olivier Patrouix, Franck Luthon, and Jean-Marc Coutellier. Dynamic background segmentation for remote reference image updating within motion detection jpeg2000. In *ICIP International Conference on Image Processing*, volume 1, 2008. [2]

[10] Liwei Zhao. *Synthesis and Acquisition of Laban Movement Analysis Qualitative Parameters for Communicative Gestures.* PhD thesis, University of Pennsylvania, 2002. [1, 2]