Pattern Recognition ■ (■■■■) ■■■–■■■

Contents lists available at SciVerse ScienceDirect

Pattern Recognition

journal homepage: www.elsevier.com/locate/pr

Exploring trace transform for robust human action recognition

Georgios Goudelis *, Konstantinos Karpouzis, Stefanos Kollias

Image, Video and Multimedia Systems Laboratory, National Technical University of Athens, 9 Heroon Politechniou Street, 15780, Athens, Greece

ARTICLE INFO

Article history:
Received 27 March 2012
Received in revised form
26 February 2013
Accepted 1 June 2013

Keywords:
Human action recognition
Motion analysis
Action classification
Trace transform

ABSTRACT

Machine based human action recognition has become very popular in the last decade. Automatic unattended surveillance systems, interactive video games, machine learning and robotics are only few of the areas that involve human action recognition. This paper examines the capability of a known transform, the so-called Trace, for human action recognition and proposes two new feature extraction methods based on the specific transform. The first method extracts Trace transforms from binarized silhouettes, representing different stages of a single action period. A final history template composed from the above transforms, represents the whole sequence containing much of the valuable spatio-temporal information contained in a human action. The second, involves Trace for the construction of a set of invariant features that represent the action sequence and can cope with variations usually appeared in video capturing. The specific method takes advantage of the natural specifications of the Trace transform, to produce noise robust features that are invariant to translation, rotation, scaling and are effective, simple and fast to create. Classification experiments performed on two well known and challenging action datasets (KTH and Weizmann) using Radial Basis Function (RBF) Kernel SVM provided very competitive results indicating the potentials of the proposed techniques.

1. Introduction

Observation and analysis of the human behavior is an open research topic for the last decade. Recognizing human actions in everyday life, is a very challenging task that can be applied in various areas such as automated crowd surveillance, shopping behavior analysis, automated sport analysis, human–computer interaction and others. One could state the problem as the ability of a system to automatically classify the action performed by a human person, given the action containing video.

Although the problem can be easily grasped, providing a solution to this problem is a daunting task that requires different approach to several sub-problems. The challenge of the task rises from various factors that influence the recognition rate.

Individuality is a major issue as the same action can be performed differently by every person. Complicated backgrounds, occlusions, illumination variations, camera stabilization and view angle are only a few of the problems that increase the complexity and create a large number of prerequisites.

If we had to classify human action recognition in different sub-categories, one could do it by taking into consideration the underlying features used to represent the various activities. As authors spot in [1], there are two main classes based on the underlying features representing activities. The most successful one is based on "dynamic features" and comprises the research object for the majority of current studies. The second one is based on "static pose based features" and provides the advantage of extracting features from still images.

A system inspired by Local Binary Patterns (LBPs) is presented in [2] that is resilient to variations in textures by comparing nearby patches, while it is silent in the absence of motion. LBP based techniques have also been proposed in [3] where the space-time volume is sliced along the three axes (x, y, t) to construct LBP histograms of the xt and yt planes. Another approach in [4] in order to capture local characteristics in optical flow, computes a variant of LBP and represents actions as strings of local atoms. In [5] another approach, inspired by biology, uses hierarchically ordered spatio-temporal feature detectors. Space time interest points are used to represent and learn human action classes in [6]. Improvement of result has reported in [7,8] where optical flow based information is combined with appearance information. In a newer study in [9], a spatiotemporal feature point detector is proposed, based on a computational model of saliency.

As mentioned above, the features used for human action recognition can be extracted either from video sequences or still images describing different static poses. The methods that use still images, are mostly silhouette based and although they do not present the accuracy of the sequence based techniques, they provide the main advantage of single frame decision extraction. Representative samples of this category are the methods presented

* Corresponding author. Tel.: +302107722491.
E-mail address: ggoudelis@image.ece.ntua.gr (G. Goudelis).

0031-3203/$ - see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.patcog.2013.06.006

in [10,11]. More specifically, in [11] behavior classification is achieved extracting eigenshapes from single silhouettes using Principal Component Analysis (PCA). Modelling of human poses from individual frames in [10],uses a bag-of-rectangles method for action categorization.

Other technique in [12] involves infrared images to extract more clear poses. In following, classification is achieved using single poses based in Histogram of Oriented Gradients (HOGs) descriptors. A type of HOG descriptors is also used in [13], on a set of predefined poses representing actions of hockey players. To better cope with articulated poses and cluttered background, authors in [14] extend HOG based descriptors and represent action classes by histograms of poses primitives. Also in contrast to other techniques that use complex action representations, authors in [15] propose a method that relies on "key pose" extraction from action sequences. The method selects the most representative and discriminative poses from a set of candidates to effectively distinguish one pose from another.

Another classification for the approaches relevant to human action recognition is attempted by authors in [16]. The difference between methods lies in the representation used by the authors. Time evolution of human silhouettes was frequently used as action description. For example, in [17] the authors proposed the representation of actions with temporal templates called Motion History (MH) and Motion Energy (ME) respectively. An extension of this study presented in [18] inspired by MH templates, introduces the Motion History Volumes as free-viewpoint representation. Working on similar direction, the authors of [19] proposed action cylinders, representing an action sequence as a generalized cylinder, while in [20] spatiotemporal volumes were generated based on a sequence of 2D contours with respect to time. These contours are the 2D projection of the points found on the outer boundary of an object performing an action in 3D. Space-time volumes shapes are also used in [21,22] based on silhouettes extracted over time.

Another recent category of techniques, also space-time oriented, is based on the analysis of the structure of local 3D patches in the action containing video [23–26]. Different local features (space-time based or not) have been combined with different machine learning techniques. Hidden Markov Models (HMMs) [27–29], Conditional Random Fields (CRFs) [30–32] and Support Vector Machines (SVMs) [33,8,34] are only a few of them.

The work introduced in this paper is an extension of the work presented in [34]. In the specific study the Radon transform was proposed for the extraction of features, capable of representing an action sequence in a form of template. Radon, which is actually a subcase of the Trace transform, has found a variety of important applications, from computerized tomography to gait recognition [35]. In this paper, we create new features examining the potentials of Trace transform for human action recognition. In the first stage of our work, we use different functionals for Trace construction, which assign different properties to a final template called *History Trace Template* (HTT). In more details, we examine different functionals of the Trace to create volume templates that each one represents a single period of an action. Radial Basis Function (RBF) kernel SVMs used for the evaluation of the technique, shows a competitive performance of 90.22% for the KTH and 93.4% for the Weizmann datasets respectively.

At a second stage, we extend further the method introducing another feature extraction technique for the production of invariant to variations features, like rotation, translation and scaling. More specifically, from each frame of the action sequence a set of Traces is calculated. Calculating different functionals with specific properties on these transforms, a set of invariant triple features is extracted. The action is finally represented by a low-dimensional vector named *History Triple Features* (HTFs) containing the most discriminant features of the sets extracted for each frame. Classification experiments using the above datasets, presented even better results of 93.14% and 95.42% respectively, indicating the potentiality of the method.

To the best of authors' knowledge, this is the first time that the Trace transform in any of its forms or its derivatives, is used for the extraction of features for human action recognition.

The rest of the paper is organized as follows. Trace transform and the theory behind it, is presented in Section 2. An overview of the proposed methods is given in Section 3. In Section 4, History Trace Template and History Triple Feature extraction techniques are described. The experimental procedure is provided in Section 5 followed by conclusion in Section 6.

2. Trace transform

Trace transform is a generalization of Radon [36] transform while at the same time Radon builds a sub-case of it. While Radon transform of an image is a 2D representation of the image in coordinates ϕ and p with the value of the integral of the image computed along the corresponding line, placed at cell (ϕ, p), Trace calculates functional T over parameter t along the line, which is not necessarily the integral. Trace transform is created by tracing an image with straight lines where certain functionals of the image function are calculated. Different transforms having different properties can be produced from the same image. The transform produced is in fact a 2-dimensional function of the parameters of each tracing line. Definition of the above parameters for an image Tracing line is given in Fig. 1. Examples of Radon and Trace transforms for different action snapshots are given in Fig. 2. In following we will give a description of the feature extraction procedure for the Trace Transform based on the theory provided in [37].

To better understand the specific transform, let us consider a linearly distorted object (rotation, translation and scaling). We could say that the object is just perceived in another coordinate system linearly distorted. This could be easier explained by letting us call the initial coordinate system of the image C_1 and the new distorted one, C_2. Let us also suppose that the distorted system can be obtained by rotating C_1 by angle $-\theta$, scaling of the axes by parameter ν and by translating with vector $(-s_0 \cos \psi_0, -s_0 \sin \psi_0)$. Suppose that there is a 2D object F which is viewed from C_1 as $F_1(x,y)$ and from C_2 as $F_2(\tilde{x},\tilde{y})$. $F_2(\tilde{x},\tilde{y})$ can be considered as an image constructed from $F_1(x,y)$ by rotation by θ, scaling by ν^{-1}, and shifting by $(s_0 \cos \psi_0, s_0 \sin \psi_0)$.

A linearly transformed image is actually transferred along lines of another coordinate system, as the straight lines in the new coordinate system also appear as straight lines. The parameters of a line in C_2 parameterized by (ϕ, p, t) in the old system C_1, are

$$\phi_{old} = \phi - \theta \tag{1}$$

$$p_{old} = \nu[p - s_0 \cos(\psi_0 - \phi)] \tag{2}$$

$$t_{old} = \nu[t - s_0 \sin(\psi_0 - \phi)]. \tag{3}$$

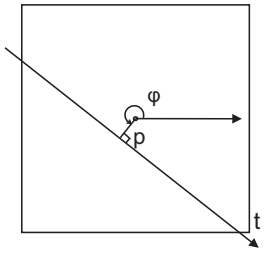

Fig. 1. Definition of the parameters of an image tracing line.

Please cite this article as: G. Goudelis, et al., Exploring trace transform for robust human action recognition, Pattern Recognition (2013), http://dx.doi.org/10.1016/j.patcog.2013.06.006

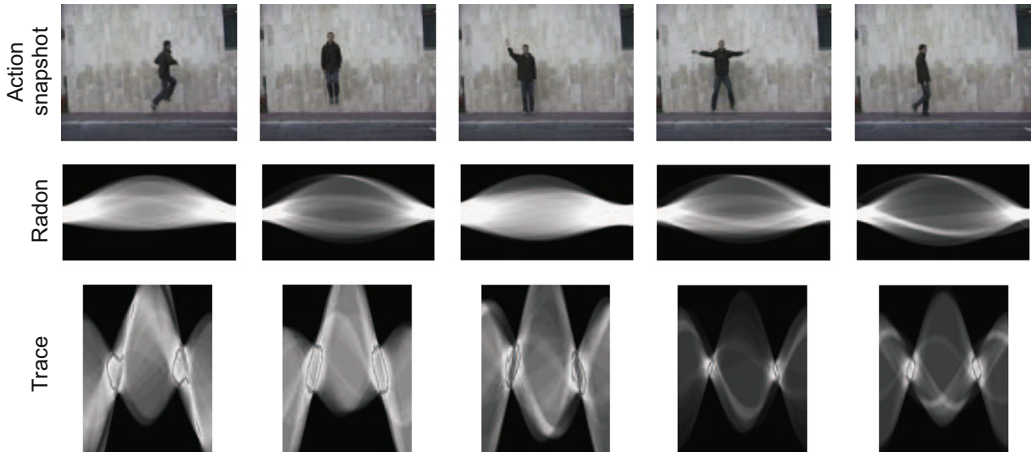

Fig. 2. Examples of Radon and Trace transforms created from the silhouettes of different action snapshots taken from Weizmann database.

Let us denote Λ as a set of lines that scan an image in all directions. The Trace transform is a function g defined on the specific set with the help of a functional T of the image function, when it is considered as a function of a variable t. Functional T is called *Trace functional*. If $L(C_1;\phi,p,t)$ is a line in coordinate system C_1, then

$$g(F;C_1;\phi,p)=T(F(C_1;\phi,p,t)), \tag{4}$$

where $F(C_1;\phi,p,t)$ means the values of the image function along a selected line. Taking this functional, variable t is eliminated. This results in a two-dimensional function of variables ϕ and p. The new function is also an image defined on Λ.

As it is described in [37], using two more functionals assigned to letters P and Φ, a *triple feature* can be defined. Where P is called diametrical and Φ is called circus functional respectively. P is a functional of the Trace transform function, when it is considered as a functional operating on the orientation variable after the previous two operations have been performed. Thus, the triple feature Π is defined as

$$\Pi(F,C_1)=\Phi(P(T(F(C_1;\phi,p,t)))). \tag{5}$$

At this point, the three functionals must be chosen. In following, *invariant* and *sensitive* to displacement functionals are used which for simplicity, will be called "invariant" and "sensitive" respectively. A functional Ξ of a function $\xi(x)$ is *invariant* if

$$\Xi(\xi(x+b))=\Xi(\xi(x)) \quad \forall b\in\Re \quad (I_1).$$

The following properties should characterize an invariant functional:

- Scaling the independent variable by α, scales the result by some factor, $a(\alpha)$

  $$\Xi(\xi(\alpha x))=a(\alpha)\Xi(\xi(x)) \quad \forall\alpha>0 \quad (i_i).$$

- Scaling the function by c scales the result by some factor, $\gamma(c)$

  $$\Xi(c\xi(x))=\gamma(c)\Xi(\xi(x)) \quad \forall c>0 \quad (i_2).$$

It has been shown that one can write

$$a(\alpha)=\alpha^{k_\Xi} \quad \text{and} \quad \gamma(c)=c^{\lambda_\Xi}, \tag{6}$$

where parameters k_Ξ and λ_Ξ characterize functional Ξ.

Functionals with the following properties are required: applied on a 2π periodic function u, the result produced should be the same with the one that would be produced if the functional, were to be applied to the original function minus its first harmonic $u^{(1)}$, denoted by $u^{\perp}\equiv u-u^{(1)}$

$$Z(u)=Z(u^{(\perp)}) \quad (si_1).$$

A functional Z is called *sensitive* if

$$Z(\zeta(x+b))=Z(\zeta(x))-b \quad \forall b\in\Re \quad (S_1).$$

A sensitive functional of a periodic function is defined as follows: Let r be the period of the function in which Z is defined. A function is called *r-sensitive* if

$$Z(\zeta(x+b))=Z(\zeta(x))-b_{(mod\ r)} \quad \forall b\in\Re \quad (S_2).$$

The following properties may also apply to a sensitive functional:

- Scaling the independent variable scales the result inversely

  $$Z(\zeta(\alpha x))=\frac{1}{\alpha}Z(\zeta(x)) \quad \forall\alpha>0 \quad (s_1)$$

  Combination of the above with (S_1), results to

  $$Z(\zeta(\alpha(x+b)))=\frac{1}{\alpha}Z(\zeta(x))-b \quad (s_{11})$$

  and

  $$Z(\zeta(\alpha x+b))=\frac{1}{\alpha}Z(\zeta(x))-\frac{b}{\alpha} \quad (s_{12}).$$

- Scaling the function does not change the result

  $$Z(c\zeta(x))=Z(\zeta(x)) \quad \forall c>0 \quad (s_2).$$

2.1. Invariant feature construction

Be it so that the functionals T,P and Φ are chosen to be invariant with T obeying property (i_1), P obeying properties (i_1) and (i_2) and Φ obeying property (i_2).

The way image linear distortion affects the value of the triple feature, is presented bellow. It can be observed that the triple feature of the distorted image is given by

$$\Pi(F,C_2)=\Phi(P(T(F(C_1;\phi_{old},p_{old},t_{old})))). \tag{7}$$

If we substitute from (1), (2) and (3), we obtain

$$\Pi(F,C_2)=\Phi(P(T(F(C_1;\phi-\theta,v[p-s_0\ \cos(\psi_0-\phi)],\ v[t-s_0\ \sin(\psi_0\ \sin(\psi_0-\phi)])))). \tag{8}$$

Using the invariance of T and property of (i_1) the above can be written as

$$\Pi(F,C_2)=\Phi(P(\alpha_T(v)T(F(C_1;\phi-\theta,v[p-s_0\ \cos(\psi_0-\phi)],t)))). \quad (9)$$

Due to property (i_2) obeyed by P, this is

$$\Pi(F,C_2)=\Phi(\gamma_p(\alpha_T(v))P(T(F(C_1;\phi-\theta,v[p-s_0\ \cos(\psi_0-\phi)],t)))). \quad (10)$$

From (i_1) property obeyed by P and its invariance, it results to

$$\Pi(F,C_2)=\Phi(\gamma_p(\alpha_T(v))\alpha_p(v)P(T(F(C_1;\phi-\theta,p,t)))). \quad (11)$$

If Φ is invariant and obeys property (i_2) there is

$$\Pi(F,C_2)=\gamma_\Phi(\gamma_P(\alpha_T(v)\alpha_p(v)))\Phi(P(T(F(C_1;\phi,p,t)))). \quad (12)$$

This condition can be expressed in terms of the exponents of the functionals κ and λ, to obtain

$$\Pi(F,C_2)=v^{\lambda_\phi(\kappa_T\lambda_p+\kappa_p)}\Pi(F,C_1). \quad (13)$$

So, invariance should followed by the obvious condition

$$\lambda_\phi(\kappa_T\lambda_p+\kappa_p)=0 \quad (14)$$

This condition is not necessary if there is no scale difference between objects that are to be matched, while any invariant functionals that obey the necessary properties can be used.

Choosing functional T, Φ to be invariant and functional P to be sensitive and obey property (s_{11}), Φ also obeys property si_1. So (10) no longer follows from (9). Instead we could apply property (s_{11}) of P, which results to

$$\Pi(F,C_2)=\Phi\left(\frac{1}{v}P(T(F(C_1;\phi,p,t))\right)+s_0\ \cos(\psi_0-\phi)). \quad (15)$$

Due to si_1 property of Φ, we obtain

$$\Pi(F,C_2)=\gamma_\Phi\left(\frac{1}{v}\right)\Phi(P(T(F(C_1;\phi,p,t)))) \quad (16)$$

or equivalently,

$$\Pi(F,C_2)=v^{-\lambda_\Phi}\Pi(F,C_1). \quad (17)$$

Choosing Φ so that

$$\lambda_\Phi=0, \quad (18)$$

it can be seen that the calculated triple feature is again invariant to rotation translation and scaling.

Conditions (14) and (18) are too restrictive though. The relationship between the triple features computed in the two cases, can be generalized by

$$\Pi(F,C_2)=v^{-\omega}\Pi(F,C_1), \quad (19)$$

for (13), $\omega\equiv-\lambda_\Phi(\kappa_T\lambda_P+\kappa_P)$, while for (17), $\omega\equiv\lambda_\Phi$. Since we can decide the type of functional that is to be constructed, we choose ω to be known. Thus, every triple feature computed can be normalized.

$$\Pi_{norm}(F,C_1)=\sqrt{[\omega]|\Pi(F,C_1)|}sign(\Pi(F,C_1)), \quad (20)$$

while (19) can be simplified to

$$\Pi(F,C_2)=v^{-1}\Pi_{norm}(F,C_1). \quad (21)$$

An invariant can be produced by dividing two triple features constructed in such a way.

3. Overview of the proposed system

The most common way to capture a human action is by using a standard 2D camera. Thus, the action is contained in a video sequence comprised by a number of different frames. In our scheme, we have worked on KTH [33] and Weizmann [38] databases. Both of them contain a large number of action video sequences, while they have been widely used for evaluation of human action recognition methods. Since background in all videos is uniform, we subtract it using a grassfire algorithm [39]. Silhouette extraction is a common technique in many different studies concerning observation of human dynamics [35,40]. As it is used in most of the human action algorithm approaches, we constructed the testing and training examples manually, segmenting (both in space and in time). We have also aligned the provided sequences. This way, each action sample is represented by a time-scaled video that contains one period.

3.1. History Trace Templates (HTTs)

Although the background is uniform, extracted silhouettes appear to be noisy as there is still a number of external factors (such as illumination conditions, etc.) that dramatically affect the result. To indicate the capabilities of the proposed methods we do not use a sophisticated algorithm for silhouette extraction neither any prior filtering. However, due to Trace transform specifications, the new features created, present to be robust to noise. Thus, a Trace transform is created for each silhouette. A final template named History Trace Template (HTT) that represents the entire movement is created as the result of the integration of the binary transformations to it.

In following, the final templates comprise the vectors that will train equal to the number of classes, RBF kernel SVMs. Examples of extracted silhouettes from frames of different actions and the HTTs produced for the specific videos, are illustrated in Fig. 3. Classification is achieved by measuring the distance of the test vector from the support vectors of each class. However, since the objective is to evaluate the overall performance of the new scheme, we measured the total number of correct classifications for every vector passing from each trained SVM respectively. For testing, we followed a leave-one-person-out protocol. Further details on the experimental procedure are provided in the corresponding Section 5.

3.2. History Triple Features (HTFs)

Exploring the capabilities of Trace transform we extended the method based on HTTs creating even more effective features for human action recognition. The new features consist of a set of triple features divisions and are invariant to different distortions.

For each video sequence, background and silhouettes are extracted as above from the same datasets. In this case, using a number of different functionals, a number of different transformations is calculated for each frame. From these transforms, a vector that is composed of a series of invariant features calculated for each frame of one period of an action is produced. Using Linear Discriminant Analysis (LDA) [44] to reduce dimensionality, the whole sequence is represented by a new vector named History Triple Feature (HTT) and is a set of real numbers containing important discriminant information for human action classification. A more comprehensive description of the specific feature extraction technique, is given in Section 4.2.

3.3. Ability of Trace to distinguish action classes (an intuitive illustration)

The features that arise from Trace transform may have not any physical meaning according to human perception. However, they may have the right mathematical properties which allow classification of actions under a certain group of transformations. To illustrate the ability of Trace to provide sufficient features for classification of actions and to provide an intuitive understanding of this ability, we constructed Weighted Trace Transforms (WTTs), which initially have been proposed in [43] for face recognition. We applied the same technique to HTTs, calculating the Weighted

Please cite this article as: G. Goudelis, et al., Exploring trace transform for robust human action recognition, Pattern Recognition (2013), http://dx.doi.org/10.1016/j.patcog.2013.06.006

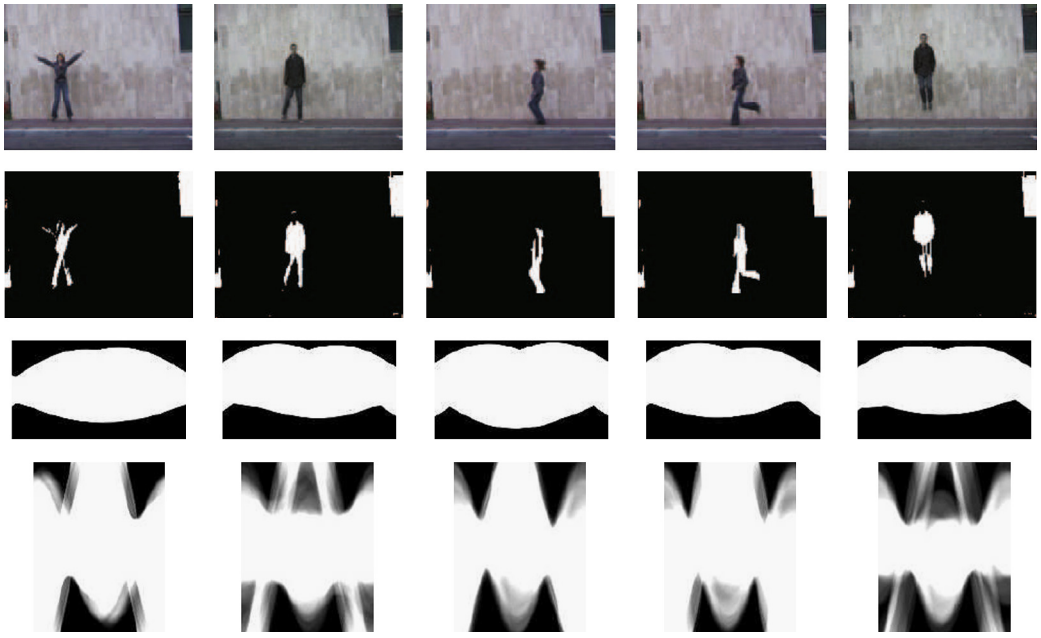

Fig. 3. Examples of History Trace Templates produced for jack, side, skip, run and pjump actions taken from Weizmann database. Second row shows extracted silhouettes for the above instances while third and fourth row show two different types of HTT, produced for each of the action videos.

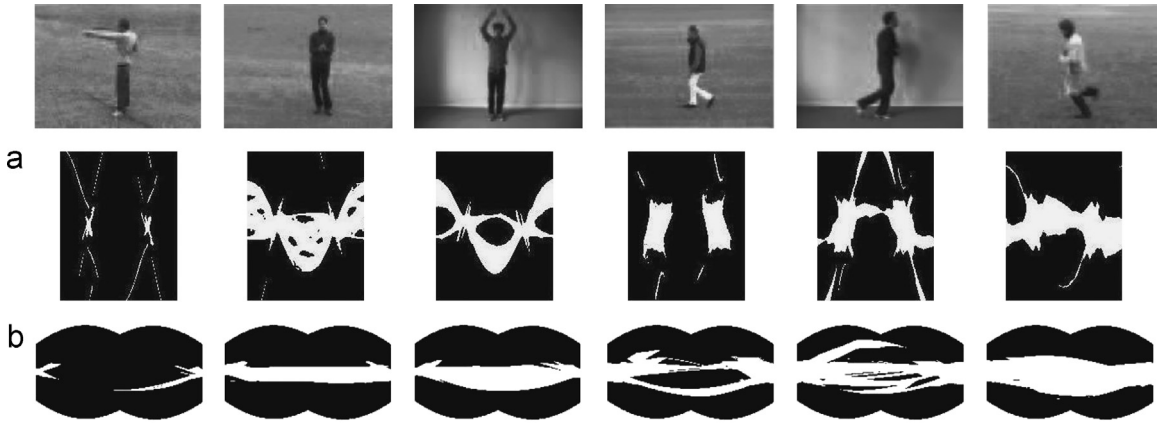

Fig. 4. Weighted History Trace templates, for all the different action classes of KTH database using two different functionals (rows (a) and (b)). Important points clearly differ from class to class.

History Trace Transforms (WHTTs) for each of the classes of the KTH database.

Every tracing line is represented by a point in the Trace representation of an image. WHTT is actually a representation of tracing lines weighted according to the role they play in the recognition of the different classes. It actually finds the features that persist in the final template (HTT) for each class, even if the action is performed by different persons or captured from different view angles. The WHTT is computed as follows:

Let D_1, D_2, D_3 be 3 training HTTs. The difference between the HTTs of the 3 actions is calculated.

$$\begin{aligned} D_1(p,\theta) &\equiv |T_1(p,\theta) - T_2(p,\theta)|, \\ D_2(p,\theta) &\equiv |T_1(p,\theta) - T_3(p,\theta)|, \\ D_3(p,\theta) &\equiv |T_2(p,\theta) - T_3(p,\theta)|, \end{aligned} \tag{22}$$

where T_i is the HTT of the ith training action and κ is a threshold. The weight matrix is defined as follows:

$$W(p,\theta) = \begin{cases} 1, & \text{if } D_1(p,\theta) \le \kappa \text{ and } D_2(p,\theta) \le \kappa \text{ and } D_3(p,\theta) \le \kappa \\ 0, & \text{otherwise.} \end{cases} \tag{23}$$

The result is finally a new template that contains those highlighted scanning lines that produced values for the HTTs that differ from each other by only up to a certain level κ. The results of the above calculations on the different classes of KTH database are illustrated in Fig. 4. WHTTs have been calculated considering every time as training set, the set of HTT samples that constitute the corresponding action class. To demonstrate that different functionals may introduce different characteristics of an action, two different functionals have been used. The difference of the flagged points between the final templates among action classes is clearly shown.

4. Constructing Trace based features for human action sequences

It has been shown [37] that the integrals along straight lines defined in the domain of a 2D function can fully reconstruct it. As it is explained above, Trace transform is produced by tracing an image along with straight lines where certain functionals of the specific function are calculated. The result of Trace transform, is another 2D image which consists a new function that depends on

the parameters (ϕ, p) that characterize each line. Different Trace transforms can be produced using different functionals. In this work, we choose the appropriate computation of the corresponding Trace functionals so that we take advantage of noise robustness of Trace and invariability to translation, and scaling.

Let $f(x,y)$ be a 2D function in the Euclidean plane $\mathbf{R}^2$ taken from an action video sequence containing an extracted binary silhouette. The Trace Transform $\breve{g}_f$, is a function defined on the space of straight lines L in $\mathbf{R}^2$ by a functional along each such line. If for instance this functional limits its operation to the integration of each line, it falls to the case of continuous Radon transform of an image and is given by

$$R_f(p,\theta) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\delta(p - x\cos\theta - y\sin\theta)\,dx\,dy \tag{24}$$

where $R(p,\theta)$ is the line integral of the image along a line from $-\infty$ to ∞. p and θ are the parameters that define the position of the line. So, $R_f(p,\theta)$ is the result of the integration of f over the line $p = x\cos\theta + y\sin\theta$. The reference point is defined as the center of the silhouette.

As human actions are in fact spatio-temporal volumes, the aim is to represent as much of the dynamic and the structural information of the action as possible. At this point, Trace transform shows a great suitability for this task. It transforms 2-dimensional images with lines into a domain of possible line parameters, where each line in the image will give a peak positioned at the corresponding line parameters. When Trace transform is calculated with respect to the center of the silhouette, specific coefficients will have capture much of the energy of the silhouette. These coefficients will vary during time and will provide great differences from one action to another for the same time-frame.

4.1. Constructing HTTs

Besides structural information, in order to also capture the temporal information included in a movement, we propose the construction of *History Trace Template*. This template is actually a continuous transform in the temporal direction of a sequence. Let $f(p,\vartheta,t)$ be a human action sequence. If $\breve{g}_n(p,\theta)$ is the Trace transform of a silhouette $s_n(p,\theta)$, for the n frame where $n = 1\ldots N$, then the History Trace Template for the action sequence will be given from

$$T_N(p,\theta) = \sum_{n=1}^{N} \breve{g}_n(p,\theta). \tag{25}$$

This way the resulting features will be a function of multiple significant distinctions contained in multiple transforms produced for every action period respectively. As mentioned above, in our work all action periods have been timescaled to the same number of frames N. Fig. 5 shows the transformations for each extracted silhouette received from one walking period. The final HTT is shown on the bottom side of the figure. For the experimental procedure, we have calculated and tested a number of Trace transforms using different functionals. The exact forms of the above transforms are provided in Table 1.

4.2. Constructing HTFs

In this section we introduce a novel human action representation using features derived from the Trace transform, hereafter simply called *History Triple Features* (HTFs). The Trace transform is a global transform that can be applied to full images. It is known to be able to pick up shape as well as texture characteristics of the

Table 1
Different functionals calculated for the experimental procedure.

Trace Transform	Functional
1	$T(f(x)) = \int_{[0,\infty]} rf(r)\,dr$ where $r = x - c$ and $c = median_x\{x, f(x)\}$
2	$T(f(x)) = \int_{[0,\infty]} r^2 f(r)\,dr$ where $r = x - c$ and $c = median_x\{x, f(x)\}$
3	$T(f(x)) = median_{r\geq 0}\{f(r), (f(r))^{1/2}\}$ where $r = x - c$ and $c = median_x\{x, f(x)\}$
4	$T(f(x)) = median_{r\geq 0}\{rf(r), (f(r))^{1/2}\}$ where $r = x - c$ and $c = median_x\{x, f(x)\}$
5	$T(f(x)) = \int_{[0,\infty]} e^{ik\log r} r^p f(r)\,dr, (p = 0.5, k = 4)$ where $r = x - c$ and $c = median_x\{x, (f(x))^{1/2}\}$
6	$T(f(x)) = \int_{[0,\infty]} e^{ik\log r} r^p f(r)\,dr, (p = 0, k = 3)$ where $r = x - c$ and $c = median_x\{x, (f(x))^{1/2}\}$
7	$T(f(x)) = \int_{[0,\infty]} e^{ik\log r} r^p f(r)\,dr, (p = 1, k = 5)$ where $r = x - c$ and $c = median_x\{x, (f(x))^{1/2}\}$

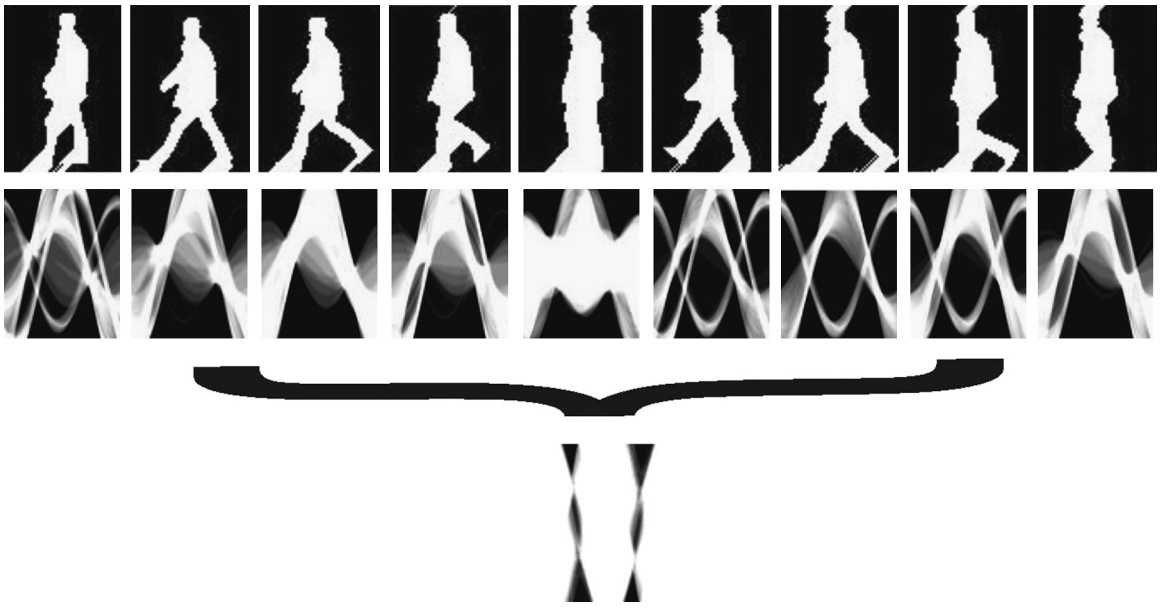

Fig. 5. Extracted silhouettes and Trace transforms for one walking period. History Trace Template is shown on the bottom.

Please cite this article as: G. Goudelis, et al., Exploring trace transform for robust human action recognition, Pattern Recognition (2013), http://dx.doi.org/10.1016/j.patcog.2013.06.006

object it is used to describe and offers an alternative representation of an image of it [43].

We presented above, how Trace can be used to represent a whole action sequence. However, the above representation can only be used in the case where the actions have been captured under the same conditions (view angle, scale, rotation, etc.). Since this is not very common in most of the applications that embroil human action recognition, we propose a more advanced technique that overcomes many of the above limitations. The Trace transform is a very rich representation of an image. To use it directly for recognition, one can produce a more simplified version of it.

Authors in [37] have proved that using extracted triple features, robust features for the classification of different but very similar to each other image classes (e.g. different kind of fishes) can be produced. In Section 2 we presented the theory behind the construction of triple features. In following, we demonstrate the construction of the proposed HTTs.

Having the extracted silhouettes, we first transform the silhouette containing image space, to Trace transform space. For each frame of the sequence, a set of Trace transforms is calculated. Following the procedure described in Section 2.1 for the extraction of the triple feature, a set of such features is extracted. The ratio of a pair of such features as it has been shown, may be invariant to different kind of distortions, depended on the functionals used. These functionals may be chosen to be sensitive or relatively insensitive to the possible variations that occur in action sequences, while maintaining discriminability.

Let $f(p,\vartheta,t)$ be a human action sequence. Applying a Trace functional T, along lines tracing the n frame referring to $s_n(p,\theta)$ silhouette, where $n=1\ldots N$ and N is the number of frames, a Trace transform $g_n(p,\theta)$ is produced. Applying different Ts to every silhouette $s_n(p,\theta)$ a set of $g_{n_i}(p,\theta)$ transforms is produced. Where $i=1\ldots L$ and L is the number of transforms one chooses to calculate. For every $g_{n_i}(p,\theta)$ a set of $\Pi_{norm}(F,C)$ normalized triple features is computed.

In a simple way triple feature is constructed as follows:

(a) Trace transform is produced by applying a Trace functional T along lines tracing the image.
(b) The circus function of the image is produced by applying a diametric functional P along columns of the Trace transform.
(c) The triple feature is finally produced by applying a circus functional Φ along the string of number produced in step b.

The procedure is illustrated in Fig. 6.

Dividing all $\Pi_{norm}(F,C)$ by each other, a set of independent features is produced. So, the whole action sequence is finally represented by a vector $\mathbf{v}$ comprised by the set of all triple feature

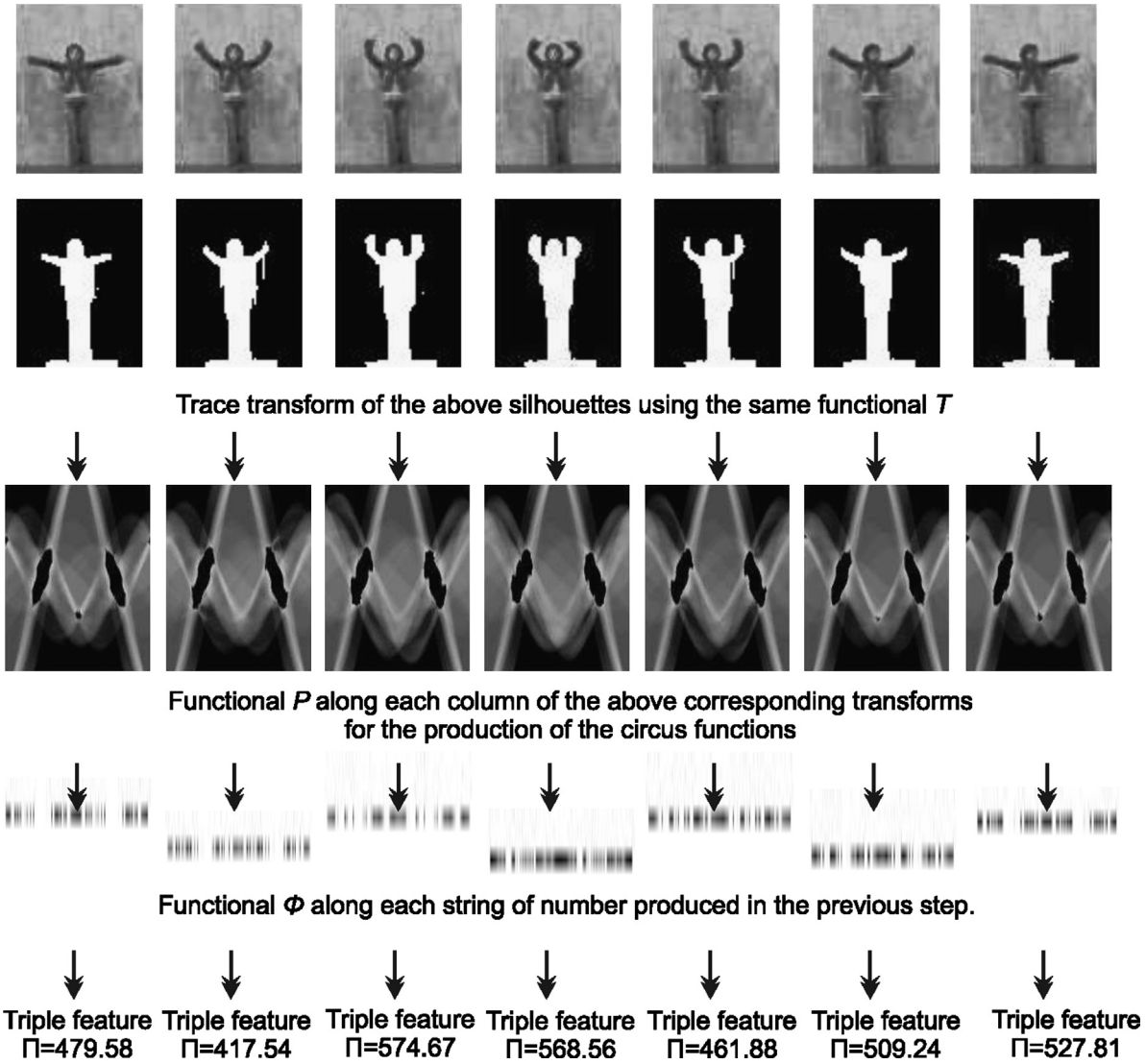

Fig. 6. Triple features extraction for one period of a wave action taken from Weizmann dataset.

ratios calculated for every frame of the action sequence.

$$\mathbf{v} = (\Pi_{rat_1}, \Pi_{rat_2}, \ldots, \Pi_{rat_{g-1}}, \Pi_{rat_g}) \quad (26)$$

where Π_{rat} is the ratio of two normalized triple features and g the number of calculated ratios.

This method allows the construction of many features easily. Supposing that one makes use of 10 functionals for each stage of the construction (e.g. 10 T functionals, 10 P functionals and 10Φ functionals) in a 10 frame action video, he may construct $10 \times 10 \times 10 \times 10 = 10\,000$ features for one sequence. As mentioned above, these numbers may have not any physical meaning according to human perception, but they may have the required mathematical properties for classification purposes.

Since the discriminatory power of the features constructed will definitely vary, a dimensionality reduction technique could provide a selection of the most discriminant features while make the problem of classification more tractable. In our scheme the HTF vectors produced as described above, become subject to LDA in order to determine an appropriate subspace that is suitable for classification. In practice we keep only a subset of the initial HTF vector that contains the most discriminant of the calculated feature capable to efficiently describe the entire action sequence.

5. Experimental results

In this section, we will present the experimental results in order to demonstrate the efficiency of the proposed schemes for human action recognition while at the same time, we will provide the experimental evaluation of the different invariant Trace functionals calculated for the construction of HTTs.

Different published methods have used different evaluation scenarios. As it is stated in [41], most of the researchers working on the field of human action recognition have evaluated their methods on the KTH [33] and Weizmann [38] datasets. However, there is not a unified standard usually followed for evaluation. The authors of the above paper also report differences up to 10.67% in results when different validation approaches are applied to the same data.

In our experiments, the leave-one-person-out cross-validation approach was used to evaluate performance. The specific protocol was chosen due to its popularity among researches. It also reconstructs the real life application needs, in the closest way. Thus, the physical dynamic behavior of an unknown subject is captured by an action recognition system and thereafter processed and compared against a pre-recorded set of data that have previously trained it. The final decision is made based on the relativity of the examined action, with one of the data that comprise the training set, according to system's set rule. Accordingly, the above protocol uses one person's samples for testing and the rest of the dataset is used for training. The procedure is repeated N times where N is the number of subjects within the dataset. Performance is reported as the average accuracy of N iterations.

The experiments were performed on an Intel Core i5 (650@3,2 GHz) processor with 4 GB RAM memory. For the experiments the KTH and the Weizmann action databases were used. Samples from the datasets used for different type of actions are illustrated in Figs. 7 and 8. The KTH video database contains six types of human actions (walking, jogging, running, boxing, hand waving and hand clapping) performed several times by 25 subjects in four different scenarios, under different illumination conditions: outdoors, outdoors with scale variation (camera zoom in and out), outdoors with different clothes and indoors. The database contains 600 sequences. All sequences were taken over homogeneous backgrounds with a static camera with 25 fps frame rate.

The Weizmann video database consists of 90 low-resolution (180×144, deinterlaced 50 fps) video sequences presenting nine different people. Each individual has performed 10 natural actions such as run, walk, skip, jumping-jack (or shortly jack), jump-forward-on-two-legs (or jump), jump-in-place-on-two-legs (or pjump), gallopsideways (or side), wave-two-hands (or wave2), wave-one-hand (or wave1), or bend.

In our experiments, the sequences have been downsampled to the spatial resolution of 160∗120 pixels and have a length of four seconds in average. The training examples were constructed by manually segmenting (both in space and in time) and aligning the available sequences. The background was removed using a grass-fire algorithm [39]. The leave-one-person-out cross-validation approach was used to test the generalization performance of the classifiers for the action recognition problem.

At this point, we should note that human action recognition is a multiclass classification problem. We cope with this, by constructing

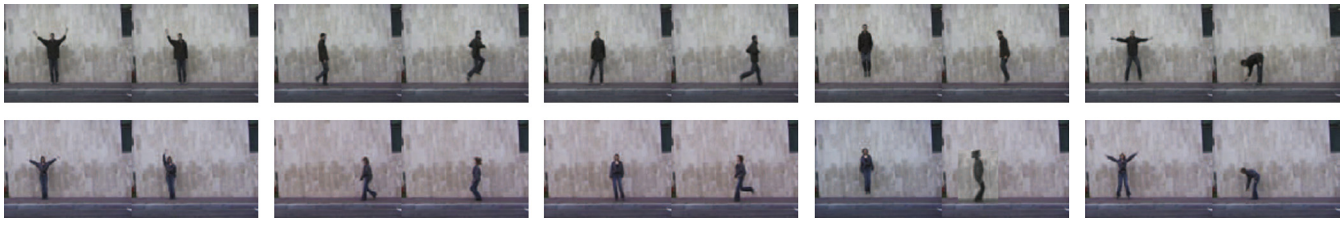

Fig. 7. Action samples from Weizmann database for wave1, wave2, walk, pjump, side, run, skip, jack, jump and bend.

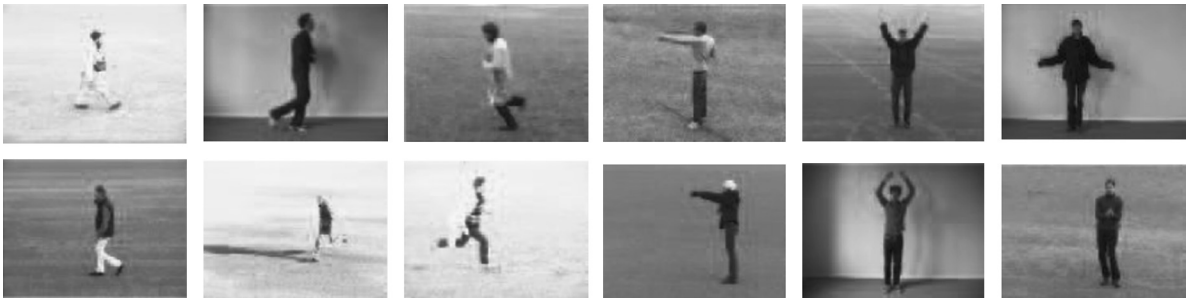

Fig. 8. Action samples from KTH database for walking, jogging, running, boxing, hand waving and hand clapping respectively.

Please cite this article as: G. Goudelis, et al., Exploring trace transform for robust human action recognition, Pattern Recognition (2013), http://dx.doi.org/10.1016/j.patcog.2013.06.006

the problem as a generalization of binary classification. More specifically, for each dataset, we trained 6 and 10 different SVMs (one for each class of the KTH and Weizmann database respectively) using an one-against-all protocol. The final decision was made by assigning each testing sample to a class $\mathcal{C}_a$, according to the distance d of the testing vector from the support vectors. Where $\mathcal{C}_a$ is the set of templates assigned to an action class (e.g. boxing). However, since we wanted to evaluate the generalization of the algorithm in a more broad way, we measured the successful binary classifications of every sample, tested on each of the different trained SVMs. This way we managed to produce $25*6*4*6 = 3600$ (persons*actions*samples per person) classifications instead of 600 for the KTH and $10*9*10 = 900$ classifications instead of 90 for the Weizmann respectively. The same procedure was followed for both feature extraction methods (HTT and HTF respectively).

The results, indicated a very competitive classification rate for both techniques. However, as it was expected, HTFs performed much better since they have been designed to be invariant to different variations. More specifically, the recognition rate using HTTs was 90.22% and 93.4% for each of the two datasets while HTFs indicated a rate of 93.14% and 95.42% for the KTH and the Weizmann databases respectively. We should also note that the extracted silhouettes used were very noisy and no prior filtering had been applied to them. The overall (for all classes) percentages produced for the different functionals of Table 1 that used for the construction of HTTs, are provided in Table 2.

It is also interesting to mention that both feature extraction methods performs quite fast. For 25 iterations (testing all samples of KTH), training included, HTTs technique required 6 min while each sample was tested within 0.01 s. The same testing procedure for the HTFs technique required 2.5 min while each sample was tested within 0.005 s. However for HTFs, testing time as well as feature extraction time, is proportional and inversely proportional respectively, to the number of Trace functionals T one chooses to calculate. In our experimental procedure, 9 T functionals were calculated for each frame of every video representing an action. This resulted in the production of 40 triple features per frame. This way, each sequence (all comprised of 7 frames), is initially described by a vector of 320 features. Using Linear Discriminant Analysis (LDA) to distinct the most discriminant of them, resulted in a vector of 31 features. Thus, every action sequence was finally represented by a 31×1 vector $\mathbf{v}$. The time required for the calculation of HTFs for one action was ≈2 s.

A comparison of the proposed techniques with other published works for the same databases are given in Tables 3 and 4 for the KTH and the Weizmann databases respectively. At this point we should note that the results illustrated are not the optimum for HTF technique. Calculating more features and/or adapting more suitable functionals for the calculation of the final HTF vector that represents a sequence may dramatically increase the results. The purpose of this work is not to present a consummate human action recognition system that gets ahead of the competition. Our aim is mainly to examine the capabilities of Trace transform for the specific task and to propose a novel feature extraction technique based on Trace that is suitable for human action recognition, while at the same time overcomes common problems such as zoom-in zoom-out and unstable video captures.

Although HTT method indicated to be able to effectively distinguish action classes, reveals some limitations when it comes to video capture variations and would probably suit more environmentally controlled applications. On the other hand, HTF indicated great potentials for the specific task as it does not only performed well, but theoretically the limitation of calculating suitable functionals is endless. For real life applications, functionals could be calculated to suit specified requirements increasing the performance of a dedicated to action recognition system. Applications such as human–computer interaction and games, could greatly benefit from a conscientious design. Outdoor applications such as surveillance systems and automated sport analyzers, could also benefit from a more generalized design that will cover specific requirements while at the same time will overcome many limitations arising from common video capturing variations.

Table 2
Results produced by calculating HTT, testing different functionals on the two different datasets.

Trace Transform	Results (%) on KTH	Results (%) on Weizmann
Radon	87.7	91.11
1	89.82	92.20
2	88.41	92.00
3	86.66	90.52
4	88.00	93.11
5	89.82	92.20
6	89.82	92.20
7	90.22	93.41

Table 3
Classification percentages (%) achieved by different published methods on KTH database.

Method	Average accuracy (%)	Classifier
Wong end Cipolla [42]	86.50	SVM
Sun et al. [45]	94.00	SVM
Liu end Shah [46]	94.16	VWCcorrel
Dollar et al. [24]	81.20	NNC
Schuldt et al. [33] (reported in [9])	50.33	NNC
Rapantzikos et al. [9]	88.30	NNC
Oikonomopoulos et al. [47] (reported in [42])	74.79	NNC
Ke et al. [48]	80.90	SVM
Schuldt et al. [33]	71.70	SVM
Niebles et al. [6]	81.50	pLSA
Jiang et al. [49]	84.40	LPBOOST
Laptev et al. [8]	91.80	SVM
HTTs	90.22	SVM
HTFs	93.14	SVM

Table 4
Classification percentages (%) achieved by different published methods on Weizmann database.

Method	Average accuracy (%)	Classifier
Sun et al. [45]	97.80	SVM
Klasser et al. [50]	84.3	SVM
Jhuang et al. [5]	96.3	SVM
Thurau [51]	86.66	MOH
Thurau et al. [14]	94.40	1-NN
Niebles et al. [6]	72.8	pLSA
HTTs	93.4	SVM
HTFs	95.42	SVM

6. Conclusion

In this paper, Trace transform is examined for its capability to produce efficient features for human action recognition. More specifically, calculating Weighted Trace transforms (WTTs) we initially presented an intuitive illustration of Trace capability to distinguish action classes. In following, two new feature extraction

Please cite this article as: G. Goudelis, et al., Exploring trace transform for robust human action recognition, Pattern Recognition (2013), http://dx.doi.org/10.1016/j.patcog.2013.06.006

methods are introduced: History Trace Templates (HTTs) and History Triple Features (HTFs). The first type of features proposed, contains much of the spatio-temporal information of a human action. A template is created integrating a series of Trace transforms calculated with different functionals that provide fast recognition and noise robustness. The second type, is produced by a series of triple features extracted from multiple Trace transforms, calculated for every frame of an action sequence. Using LDA to reduce dimensionality, the whole action is finally represented by a small vector containing the most discriminant features. The vectors produced, may have not any physical meaning according to human perception, they contain though, the mathematical properties required for action classification. The features produced are calculated to be invariant to scaling, translation and rotation, while they are noise robust giving solutions to some of the most important problems in the field of action recognition.

We evaluated the effectiveness of both methods for the specific task, by calculating different functionals for HTT and HTF and testing the new features on KTH and Weizmann databases. It is worth noting, that both methods proved to be very effective for action recognition showing great noise robustness. However, HTFs, are suggested for a wider range of applications as they perform better, present invariability to different scenarios and are very robust in illumination variations, noise and scaling (zoom-in zoom-out) conditions. The method is of great potentiality as one may calculate more, different and more reliable features that can present invariability to conditions appeared in specific type of action applications.

Conflict of interest statement

None declared.

Acknowledgement

This work was co-financed by ICT-Project Siren, under contract (FP7-ICT-258453), by the European Union (European Social Fund ESF) and Greek national funds through the Operational Program “Education and Lifelong Learning” of the National Strategic Reference Framework (NSRF) - Research Funding Program: THALES. Investing in knowledge society through the European Social Fund.

References

[1] C. Thurau, V. Hlavac, Pose primitive based human action recognition in videos or still images, in: CVPR, IEEE Computer Society, Madison, USA, Omnipress, 2008, p. 8.
[2] L. Yeffet, L. Wolf, Local trinary patterns for human action recognition, in: (ICCV), IEEE, 2009.
[3] V. Kellokumpu, G. Zhao, M. Pietikainen, Human activity recognition using a dynamic texture based method, in: Proceedings of the British Machine Vision Conference (BMVC), Leeds, UK, p. 10.
[4] C. Yang, Y. Guo, H. Sawhney, R. Kumar, Learning actions using robust string kernels, in: HUMO07, Springer, 2007, pp. 313–327.
[5] H. Jhuang, T. Serre, L. Wolf, T. Poggio, A biologically inspired system for action recognition, in: ICCV, IEEE, 2007, pp. 1–8.
[6] J.C. Niebles, H. Wang, L. Fei-Fei, Unsupervised learning of human action categories using spatial-temporal words, International Journal of Computer Vision 79 (2008) 299–318.
[7] K. Schindler, L. Van Gool, Action snippets: how many frames does human action recognition require? in: CVPR08, 2008, pp. 1–8.
[8] I. Laptev, M. Marszałek, C. Schmid, B. Rozenfeld, Learning realistic human actions from movies, in: Conference on Computer Vision & Pattern Recognition (CVPR), IEEE, 2008.
[9] K. Rapantzikos, Y. Avrithis, S. Kollias, Dense saliency-based spatiotemporal feature points for action recognition, in: CVPR, 2009.
[10] R. Goldenberg, R. Kimmel, E. Rivlin, M. Rudzsky, Behavior classification by eigendecomposition of periodic motions, in: Pattern Recognition, 2005, pp. 38:1033–1043.
[11] N. Ikizler, P. Duygulu, Human action recognition using distribution of oriented rectangular patches, in: ICCV, Human Motion, 2007, pp. 271–284.
[12] L. Zhang, B. Wu, R. Nevatia, Detection and tracking of multiple humans with extensive pose articulation, in: ICCV, 2007, pp. 1–8.
[13] W.-L. Lu, J.J. Little, Simultaneous tracking and action recognition using the pca-hog descriptor, in: CRV, IEEE Computer Society, 2006, p. 6.
[14] C. Thurau, V. Hlavac, Pose primitive based human action recognition in videos or still images, in: CVPR, IEEE Computer Society, 2008, pp. 1–8.
[15] S. Baysal, M.C. Kurt, P. Duygulu, Recognizing human actions using key poses, in: International Conference on Pattern Recognition, 2010, pp. 1727–1730.
[16] I.N. Junejo, E. Dexter, I. Laptev, P. Perez, View-independent action recognition from temporal self-similarities, IEEE Transactions on Pattern Analysis and Machine Intelligence 33 (2011) 172–185.
[17] A.F. Bobick, J.W. Davis, The recognition of human movement using temporal templates, IEEE Transactions on Pattern Analysis and Machine Intelligence 23 (2001) 257–267.
[18] D. Weinland, R. Ronfard, E. Boyer, Free viewpoint action recognition using motion history volumes (November/December 2006).
[19] T. Syeda-Mahmood, A. Vasilescu, S. Sethi, Recognizing action events from multiple viewpoints, in: IEEE Workshop on Detection and Recognition of Events in Video, 2001, p. 64.
[20] A. Yilmaz, M. Shah, Actions sketch: a novel action representation, in: CVPR (1)'05, 2005, pp. 984–989.
[21] M. Grundmann, F. Meier, I. Essa, 3d shape context and distance transform for action recognition, in: 19th International Conference on Pattern Recognition, 2008. ICPR 2008, 2008, pp. 1–4.
[22] L. Gorelick, M. Blank, E. Shechtman, M. Irani, R. Basri, Actions as space-time shapes, Transactions on Pattern Analysis and Machine Intelligence 29 (12) (2007) 2247–2253.
[23] E. Shechtman, M. Irani, Space-time behavior based correlation -or- how to tell if two underlying motion fields are similar without computing them? in: In IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI), vol. 29, 2007, pp. 2045–2056.
[24] P. Dollar, V. Rabaud, G. Cottrell, S. Belongie, Behavior recognition via sparse spatio-temporal features, in: Proceedings of the 14th International Conference on Computer Communications and Networks, IEEE Computer Society, Washington, DC, USA, 2005, pp. 65–72.
[25] I. Laptev, On space-time interest points, International Journal of Computer Vision 64 (2005) 107–123.
[26] A. Gilbert, J. Illingworth, R. Bowden, Scale invariant action recognition using compound features mined from dense spatio-temporal corners, in: Proceedings of the 10th European Conference on Computer Vision: Part I, ECCV '08, Springer-Verlag, Berlin, Heidelberg, 2008, pp. 222–233.
[27] S. Ali, A. Basharat, M. Shah, Chaotic invariants for human action recognition, ICCV 206.
[28] K. Jia, D.-Y. Yeung, Human action recognition using local spatio-temporal discriminant embedding, in: IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2008, 2008, pp. 1–8.
[29] D. Weinland, E. Boyer, Action recognition using exemplar-based embedding, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Anchorage, 2008, pp. 1–7.
[30] L. Wang, D. Suter, Recognizing human activities from silhouettes: motion subspace and factorial discriminative graphical model, in: CVPR, 2007.
[31] C. Sminchisescu, A. Kanaujia, D. Metaxas, Conditional models for contextual human motion recognition, Computer Vision and Image Understanding 104 (2006) 210–220.
[32] L.P. Morency, A. Quattoni, T. Darrell, Latent-dynamic discriminative models for continuous gesture recognition, in: IEEE Conference on Computer Vision and Pattern Recognition. CVPR '07, 2007, pp. 1–8.
[33] C. Schuldt, I. Laptev, B. Caputo, Recognizing human actions: a local SVM approach, in: In Proceedings of the ICPR, 2004, pp. 32–36.
[34] G. Goudelis, K. Karpouzis, S. Kollias, Robust human action recognition using history trace templates, in: 12th International Workshop on Image Analysis for Multimedia Interactive Services (WIAMIS), Delft, The Netherlands, 13–15 April, 2011.
[35] N.V. Boulgouris, Z.X. Chi, Gait recognition using radon transform and linear discriminant analysis, IEEE Transactions on Image Processing 16 (3) (2007) 731–740.
[36] S.R. Deans, The Radon Transform and Some of Its Applications, Krieger Publishing Company, 1983.
[37] A. Kadyrov, M. Petrou, The trace transform and its applications, IEEE Transactions on Pattern Analysis and Machine Intelligence 23 (2001) 811–828.
[38] M. Blank, L. Gorelick, E. Shechtman, M. Irani, R. Basri, Actions as space-time shapes, in: The Tenth IEEE International Conference on Computer Vision (ICCV'05), 2005, pp. 1395–1402.
[39] G. Goudelis, A. Tefas, I. Pitas, Automated facial pose extraction from video sequences based on mutual information, IEEE Transactions on Circuits and Systems for Video Technology 18 (3) (2008) 418–424.
[40] I. Kotsia, I. Patras, Relative margin support tensor machines for gait and action recognition, in: CIVR, 2010, pp. 446–453.
[41] Z. Gao, M.-Y. Chen, A.G. Hauptmann, A. Cai, Comparing evaluation protocols on the kth dataset, in: Proceedings of the First International Conference on Human Behavior Understanding, HBU'10, Springer-Verlag, Berlin, Heidelberg, 2010, pp. 88–100.

Please cite this article as: G. Goudelis, et al., Exploring trace transform for robust human action recognition, Pattern Recognition (2013), http://dx.doi.org/10.1016/j.patcog.2013.06.006

[42] S.-F.Wong, R. Cipolla, Extracting spatiotemporal interest points using global information, in: IEEE International Conference on Computer Vision, 2007, pp. 1–8.
[43] S. Srisuk, M. Petrou, W. Kurutach, A. Kadyrov, A face authentication system using the trace transform, in: Pattern Analysis and Applications, September 2005, 8, pp. 50–61, issn 1433-7541.
[44] R.O. Duda, P.E. Hart, D.G. Stork, Pattern Classification, Wiley, New York, 2000.
[45] X. Sun, M. Chen, A. Hauptmann, Action recognition via local descriptors and holistic features, Computer Vision and Pattern Recognition Workshop 0 (2009) 58–65.
[46] J. Liu, M. Shah, Learning human actions via information maximization, Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on , vol., no., pp.1,8, 23-28 June 2008 http://dx.doi.org/10.1109/CVPR.2008.4587723.
[47] A. Oikonomopoulos, I. Patras, M. Pantic, Spatiotemporal salient points for visual recognition of human actions, IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics 36 (3) (2005) 710–719. http://dx.doi.org/10.1109/TSMCB.2005.861864.
[48] Y. Ke, R. Sukthankar, M. Hebert, Spatio-temporal shape and flow correlation for action recognition, in: In Seventh International Workshop on Visual Surveillance, 2007.
[49] H. Jiang, M.S. Drew, Z. nian Li, Successive convex matching for action detection, in: In Proceedings of the CVPR, Press, 2006, pp. 1646–1653.
[50] A. Kläser, M. Marszałek, C. Schmid, A spatio-temporal descriptor based on 3d-gradients, in: British Machine Vision Conference, 2008, pp. 995–1004.
[51] C. Thurau, Behavior histograms for action recognition and human detection, in: Proceedings of the Second Conference on Human Motion: Understanding, Modeling, Capture and Animation, Springer-Verlag, Berlin, Heidelberg, 2007, pp. 299–312.

Georgios Goudelis received his B.Eng. (2003) in Electronic Engineering with Medical Electronics and his M.Sc. (2004) in Electronic Engineering both from University of Kent at Canterbury. From 2004 to 2007 he has worked as a researcher with the Artificial Intelligence and Information Analysis lab of the Department of Informatics at the Aristotle University of Thessaloniki participating in several projects financed by National and European funds. Currently, he is pursuing the Ph.D. degree in artificial intelligence and information analysis from the department of Electrical and Electronic Engineering, National Technical University of Athens, Greece. He has authored and co-authored several journal papers and international conferences and contributed in one book chapter in his area of expertise. His current research interests include, statistical pattern recognition especially for human actions localization and recognition, computational intelligence and computer vision.

Konstantinos Karpouzis graduated from the School of Electrical and Computer Engineering, of the National Technical University of Athens in 1998 and received his Ph.D. degree in 2001 from the same University. His current research interests lie in the areas of natural human computer interaction, serious games, emotion understanding and affective and assistive computing. He has published more than 110 papers in international journals and proceedings of international conferences and has contributed to the 'Humaine Handbook on Emotion research' and the 'Blueprint for an affectively competent agent'. Since 1995 he has participated in more than twelve research projects at Greek and European level; most notably the Humaine Network of Excellence, in the field of mapping signals to signs of emotion and the FP7 STREP Siren on serious games for conflict resolution, where he is the Technical Manager.

Stefanos Kollias received the Diploma degree in Electrical Engineering from the National Technical University of Athens (NTUA) in 1979, the M.Sc. degree in Communication Engineering from the University of Manchester (UMIST), U.K., in 1980, and the Ph.D. degree in Signal Processing from the Computer Science Division of NTUA in 1984. In 1982 he received a ComSoc Scholarship from the IEEE Communications Society. From 1987 to 1988, he was a Visiting Research Scientist in the Department of Electrical Engineering and the Center for Telecommunications Research, Columbia University, New York, U.S.A. Since 1997 he is Professor of NTUA and Director of IVML. He is member of the Executive Committee of the European Neural Network Society.

Please cite this article as: G. Goudelis, et al., Exploring trace transform for robust human action recognition, Pattern Recognition (2013), http://dx.doi.org/10.1016/j.patcog.2013.06.006