A Hybrid Object Tracking for Hand Gesture Approach based on MS-MD and its Application

Amir Hooshang Mazinan*
Department of Engineering, South Tehran Branch, Islamic Azad University, Tehran, Iran
ahmazinan@gmail.com
Jalal Hassanian
Department of Engineering, South Tehran Branch, Islamic Azad University, Tehran, Iran
jalalhasanian@gmail.com

Received: 12/Dec/2014 Revised: 16/Aug/2015 Accepted: 19/Aug/2015

Abstract

In the research presented here, the hand gesture recognition is considered in American Sign Language to propose a hybrid approach, which is now organized based on the integration of the mean shift (MS), and the motion information (MD), entitled MS-MD approach. This is in fact proposed to track and recognize the hand gesture, in an effective manner, along with the corresponding potential benchmarks. The investigated results are synchronously acquired in line with these well-known techniques in the area of object tracking to modify those obtained from the traditional ones. The MS scheme is capable of tracking the objects based on its detailed objects, so we have to specify ones, as long as the MD scheme is not realized. In the proposed approach, the advantages of two algorithms are efficiently used, in a synchronous manner, to outperform the hand tracking performance. In the first step, the MD scheme is applied to remove a number of parts without area motion, and subsequently the MS scheme is accurately realized to deal with hand tracking. Subsequently, the present approach is considered to eliminate the weakness of the traditional methods, which are only organized in association with the MS scheme. The results are all carried out on Boston-104 database, where the hand gesture is tracked, in a better form, with respect to the previous existing ones.

Keywords: Hand Tracking; Motion Detection; Mean Shift; Hand Gesture Recognition; American Sign Language.

1. Introduction

Due to the fact that the recognition and tracking of hand gesture are so applicable in both academic and real environments, the new, constructive, impressive and efficient insights in this area can always be appreciated. It means that state-of-the-art in the area of object tracking with a focus on hand tracking is a challenging issue among the experts of the present field. It is known as a crucial and basic ingredient computer vision. Humans can simply recognize and track an object, in an immediate manner, even in the presence of high clutter, occlusion and non-linear variations in the background as well as in the shape, direction or even the size of the object. However, hand tracking can be taken into real consideration as a difficult task for an intelligence-based machine. Tracking for a machine can be defined to find the object states. It is included by position, scale, velocity, feature selecting and many other important parameters that are obtained through a series of images, so object tracking is processed, at each incoming frame, to obtain the weighting coefficients of the entire image. Therefore, it is necessary to specify the object that is recognized in the hand gesture tracking, while a specific one is considered, as the desired object. Many solutions are now proposed to deal with hands motion that they are all used some features of objects to be obtained such as colors, area motion, texture and so on. Firstly, their method converted color frames into gray level images, and then a

kernel function is employed. Furthermore, weights of pixels are obtained for each frame. Their proposed method offers several advantages. For instance, it can be known, as a so resistant approach, against environmental difficulties such as partial occlusion, blurring via camera shaking, deformation of object position and any other sorts of translation. This is due to employing color information as feature vectors in the proposed technique.

Literature's survey in this area could now be presented by considering the work of Nguyen et al., at first, which propose tracking and recognition for facial expressions in American Sign Language; ASL. This work describes towards recognizing facial expressions that are used in sign language communication [1]. Munib et al. suggest the recognition based Hough transform and neural networks. The outcome investigated in this research aims to develop a system for automatic translation of static gestures regarding the alphabets and signs [2]. Coleca et al. describes self-organizing maps for hand and full body tracking [3], where Lee et al. present hand rotation and various grasping gestures tracking from the IR camera via extended cylindrical manifold embedding [4]. Moreover, Morshidi et al. suggest hand tracking through gravity optimized particle filter [5], while Huang et al. research is given in kinematic property of object motion conditions and eye-hand synergy during manual tracking [6]. Also,

* Corresponding Author

there is another novel research, presented by Inoue et al. [7], while robust surface tracking in range image sequences is investigated by Husain et al. [8]. Moreover,

electronics control is proposed by Premaratne et al. This work is based on the well-known wave controller technology [9]. Hereinafter, Ge et al. research is in hand gesture recognition and tracking via the distributed linear embedding [10], once González-Ortega work is to realize real-time hands, facial features detection and tracking with its application to cognitive rehabilitation tests monitoring [11]. Cui et al. propose model-based visual hand posture tracking for guiding a dexterous robotic hand [12], while Kodagoda et al. present simultaneous tracking and motion pattern learning [13]. Moreover, Guazzini et al. research is to present cognitive dissonance and social influence effects on preference judgments for eye tracking based system for their automatic assessment, as another work in this area [14]. Kılıboz et al. present a hand gesture recognition technique for human–computer interaction [15], where Li et al. propose feature learning based on SAE–PCA network for human gesture recognition in RGBD images [16]. Rautaray et al. suggest vision based hand gesture recognition for human computer interaction [17], while K. Li et al. exploit object tracking method based on mean shift and particle filter [18]. Zhang et al. introduce real-time hand gesture tracking based on particle filtering [19], where Hsia et al. present moving target tracking based on camshaft approach [20]. Shan et al. discuss real-time hand tracking through a mean shift embedded particle filter [21]. Finally, Jacob et al. exploit context-based hand gesture recognition for the operating room [22], where Kong et al. focus on independent continuous sign language recognition, as well [23].

The key goal of the approach proposed here is to present a tracking system to be applicable in the area of American Sign Language with acceptable accuracy and the appropriate time consumed, where the outcomes could be competitive with respect to the potential benchmark. Moreover, the bottlenecks of the present approach are that the limitations in choosing the objects to be tracked, the appropriate thresholds to be initialized and finally the hardware performance are existed.

The rest of the present manuscript is organized as follows: in Section 2, the proposed approach is presented. The experiments and results are shown in Section 3 and finally Section 4 draws the conclusions.

2. The Proposed Approach

The proposed approach, as an integration of two well-known algorithms including the mean shift (MS) scheme and the motion detection (MD) scheme is now presented, as a hybrid hand gesture tracking, which is working in the presence of occlusion problem, as soon as hand may be put on the face, especially. In this approach, the hand model is first determined by an authorized user, as a main tracking object, in the first frame. Then, motion range is obtained via the MD scheme to eliminate the move less parts, so having the less pixels results in minimum time for hand tracking. Hand color feature is extracted from this specific space by the MS scheme and a number of pixels are determined in the same space. Subsequently, pixels that having less intensity as compared to the hand color could be eliminated. Remained pixels are weighted in accordance with the distance from the hand center and, in the next frames, the pixels weight of the same hand center are tracked by the MS scheme. The schematic diagram of the proposed approach is sketched in Fig. 1.

Regarding the present approach, it should be noted that the main contribution of this research is to realize a new hybrid of the well-known MS scheme in association with the MD scheme, as long as its application is efficient in the specific application of object tracking. Furthermore, it is needed to note that the performance of this algorithm depends on predefined thresholds. It means that the appropriate value for this parameter needs to be first initiated, in its correct form, prior to running the program in the present research. It should be noted that the number of identified objects may correspondingly be increased, while the value of this parameter is initiated lower than its nominal one. Subsequently, by choosing the same parameter higher than its nominal one, the number of identified objects may correspondingly be decreased. In both cases, the approach performance may be poor by reaching object tracking errors. In fact, the contribution of the present research has to be twofold. First of all, the main approach of the research is recently considered in the area of paper's topic. Second, the results are somehow competitive w. r. t. the other related potential ones, as considered in the proceeding subsections.

2.1 The MD Scheme Realization

The MD scheme is a well-known approach in the area of object tracking, which is in fact faster with respect to the MS scheme-based tracker. Also, the present MD scheme is the simplest way in case of the three tasks including detection, estimation and segmentation, respectively. It should be noted that its goal is to identify which image points, or more generally which regions of the image have moved between two instants of time. The MD scheme is realized to compare the current frame with the previous one. The results are useful in video compression, as long as it is needed to estimate changes and to write these changes instead of frames. This algorithm presents an image with white pixels (motion level) in the place, where the current frame is different from the previous one. It is already possible to count these pixels and if the amount of these pixels may be greater than a predefined threshold level, this is produced as a motion event. The motion detection is calculated though

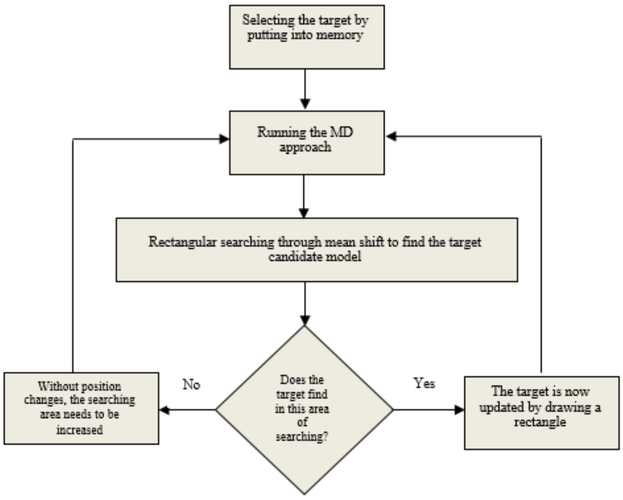

Fig.1. The schematic of the proposed approach.

the distance in the luminance space between the current image $I_k(x)$ and the last aligned image $\hat{I}_{k-1}(x)$, by obtaining the difference image $DM_k(x)$, defined as

$$DM_k(x) = \begin{cases} \text{mIncrement}; \left|I_k(x) - \hat{I}_{k-1}(x)\right| > T_m \\ 0; \qquad \text{Otherwise} \end{cases} \quad (1)$$

where mIncrement corresponds to a factor of incrementing in the motion and T_m corresponds to a motion threshold. $DM_k(x)$ contains the initial set of points that are taken as a candidate to be belonged to the moving visual object. In order to consolidate the blobs to be detected, a 3×3 morphological closing is applied to $DM_k(x)$. Isolated detected moving pixels are discarded, when applying to a 3×3 morphological opening. The representation of the motion history image $MH_k(x)$ is then updated by multiplying the last motion history representation $MH_{k-1}(x)$ with a decay factor and by adding the difference image $DM_k(x)$, where it could be written as $MH_k(x) = MH_{k-1}(x)\,.\,\text{decay factor} + DM_k(x)$. Finally, all pixels of $MH_k(x)$, whose luminance is over the MD scheme threshold are considered as pixels in motion. These pixels generate the detection image $DH_k(x)$, defined as

$$DH_k(x) = \begin{cases} 1; MH_k(x) > T_k \\ 0; \qquad \text{Otherwise} \end{cases} \quad (2)$$

A 3×3 morphological closing is now applied to the detection image $DH_k(x)$, followed by a 3×3 morphological opening.

2.2 The MS Scheme Realization

2.2.1 Target Representation

The MS scheme is realized to characterize the object, at first, while a feature space needs to be chosen. The reference object model is represented by its pdf q in the feature space. In the subsequent frame, an object candidate must be defined in location y and is characterized by the pdf $\hat{p}(y)$. Both pdfs are to be estimated from the data. To satisfy the low computational cost, it is imposed by real-time processing discrete densities, i.e., m -bin histograms. Thus, there are the object model and its candidate as

$$\begin{cases} \hat{q} = \{\hat{q}_u\}_{u=1,2,\cdots,m};\ \sum_{u=1}^{m} \hat{q}_u = 1 \\ \hat{p} = \{\hat{p}_u(y)\}_{u=1,2,\cdots,m};\ \sum_{u=1}^{m} \hat{p}_u = 1 \end{cases} \quad (3)$$

Now, the histogram is not the best nonparametric density estimation, but it suffices for our purposes. Other discrete density estimation can also be employed, where it denotes by $\hat{p} = \rho[\hat{p}(y), \hat{q}]$as a similarity function between $\hat{p}(y)$ and $\hat{q}$. The function $\hat{p}(y)$ plays the important role of likelihood and its local maximum in the image indicating the presence of objects in the second frame that having representations, which are similar to $\hat{q}$, defined in the first frame. To find the maximum of such functions, gradient-based optimization procedures are difficult to apply and only an expensive exhaustive search can be used. We regularize the similarity function by masking the objects with an isotropic kernel in the spatial domain. As long as the kernel weights, carrying continuous spatial information, are used in defining the feature space representations, $\hat{p}(y)$ becomes a smooth function in y.

2.2.2 Target Model

An object is represented by an ellipsoidal or rectangular region in the image. To eliminate the influence of different object dimensions, all the objects are first normalized to be taken in a unit circle. This is achieved by independently rescaling with the row and column dimensions along with h_x and h_y, respectively. Let us note $\{x_i^*\}_{u=1,2,\cdots,n}$ as the normalized pixel locations in the region, namely the object model. The region is centered at zero. An isotropic kernel, with a convex and monotonic decreasing kernel profile $k(x)$, assigns smaller weights to pixels, which are farther from the center. The function $b: R^2 \to \{1,2,\cdots,m\}$ associates to be the pixel at location x_i^* and also the index $b(x_i^*)$ of its bin in the quantized feature space. The probability of the feature $u = 1,2,\cdots,m$ in the object model is then calculated as $\hat{q}_u = C\sum_{i=1}^{n} k(\|x_i^*\|^2)\delta[b(x_i^*) - u]$, where $k(x)$ is the Kronecker delta function. The normalization constant $C = \frac{1}{\sum_{i=1}^{n} k(\|x_i^*\|^2)}$ is derived by imposing the condition $\sum_{u=1}^{m} \hat{q}_u = 1$, where the summation of delta functions for $u = 1,2,\cdots,m$ is equal to be one.

2.2.3 Target Candidates

Let us note $\{x_i\}_{u=1,2,\cdots,n_k}$ as the normalized pixel locations of the object candidate, centered at y in the current frame. Using the same kernel profile $k(x)$, with bandwidth h, the probability of the feature $u = 1,2,\cdots,m$ in the object candidate could be given by $\hat{p}_u(y) =$

$$C_h \sum_{i=1}^{n_k} k\left(\left\|\frac{y-x_i}{h}\right\|^2\right)\delta[b(x_i) - u] \quad (4)$$

where $C = \frac{1}{\sum_{i=1}^{n_k} k(\left\|\frac{y-x_i}{h}\right\|^2)}$ is taken. Note that C_h cannot depend on y, as long as the pixel locations x_i are organized in a regular lattice and y is taken as one of the lattice nodes. Therefore, C_h can be recalculated for a given kernel and different values of h, while the bandwidth h defines the scale of the object candidate, i.e., the number of pixels, considered in the localization process.

2.2.4 Similarity Function Smoothness

The similarity function inherits the properties of the kernel profile $k(x)$, while the object model and candidate are represented. The employed object representations are not restricting the way, which similarity is measured and various functions can be used for ρ.

2.2.5 Metric based on Bhattacharyya Coefficient

The similarity function illustrates a distance among object model and its candidates. To accommodate comparisons in the various objects, this should have a metric structure. The distance between two discrete distributions is now given by $d(y) = \sqrt{1 - \rho[\hat{p}(y), \hat{q}]}$ where $\hat{p}(y) = \rho[\hat{p}(y), \hat{q}] = \sum_{u=1}^{m} \sqrt{\hat{p}(y)\hat{q}}$ is taken. It could be noted that the cosine of the angle between the m-dimensional unit vectors $(\sqrt{\hat{p}_1}, \sqrt{\hat{p}_2}, \cdots, \sqrt{\hat{p}_m})^T$ and also $(\sqrt{\hat{q}_1}, \sqrt{\hat{q}_2}, \cdots, \sqrt{\hat{q}_m})^T$ are given. Note that the L_P histogram metrics including histogram intersection cannot satisfy the conditions $\sum_{u=1}^{m} \hat{p}_u = 1$ and$\sum_{u=1}^{m} \hat{q}_u = 1$.

2.2.6 Target Localization

In order to find the location, corresponding to the object, in the current frame, the distance should be minimized as a function of y. The localization procedure starts from the position of the object in the previous frame, i.e. the model and by searching in the neighborhood. Since the distance function is smooth, the procedure uses gradient information, provided by the MS scheme vector. More involved optimizations can be applied. Color information could be chosen as the object feature; however, the same framework may be used for texture and edges or any combination of them. In the sequel, it is assumed that the following information is available, i.e. (1) detection and localization in the initial frame of the objects to track (object models); (2) periodic analysis of each object to account for possible updates in case of the object models, due to significant changes in color.

2.2.7 Distance Minimization

Minimizing the distance is equivalent to maximize the Bhattacharyya coefficient $\hat{p}(y)$. It should be noted that the search for the new object location, in the current frame, starts at the location $\hat{y}_0$ of the object in the previous frame. It is obvious that the probabilities $\{\hat{p}_u(\hat{y}_0)\}_{u=1,2,\cdots,m}$ of the object candidate at location $\hat{y}_0$, in the current frame, need to be computed, firstly. Using Taylor expansion around the values $\hat{p}_u(\hat{y}_0)$, the linear approximation of the Bhattacharyya coefficient could be acquired after some manipulations as

$$\rho[\hat{p}(y), \hat{q}] \approx \frac{1}{2}\sum_{u=1}^{m} \frac{\hat{q}_u}{\hat{p}_u(\hat{y}_0)} + \frac{1}{2}\sum_{u=1}^{m} \sqrt{\hat{p}_u(\hat{y}_0)\hat{q}_u} \quad (5)$$

The approximation is satisfactory, as long as the object candidate $\{\hat{p}_u(\hat{y}_0)\}_{u=1,2,\cdots,m}$ cannot change, drastically, from the initial $\{\hat{p}_u(\hat{y}_0)\}_{u=1,2,\cdots,m}$. The condition $\hat{p}_u(\hat{y}_0) > 0$ or some small threshold for all $u = 1,2,\cdots,m$ can always be enforced to avoid using the feature values in violation. Recalling the outcomes could be used in

$$\rho[\hat{p}(y), \hat{q}] \approx \frac{C_h}{2}\sum_{u=1}^{n_k} w_i\, k(\left\|\frac{y-x_i}{h}\right\|^2) + \frac{1}{2}\sum_{u=1}^{m} \sqrt{\hat{p}_u(\hat{y}_0)\hat{q}_u} \quad (6)$$

where $w_i = \sum_{u=1}^{m} \sqrt{\frac{\hat{q}_u}{\hat{p}_u(\hat{y}_0)}}\,\delta[b(x_i) - u]$. The mode of this density in the local neighborhood is the sought maximum which can be found by employing the MS scheme. In this procedure, the kernel is recursively moved from the current location $\hat{y}_0$ to the new location $\hat{y}_1$ in accordance with the following relation

$$\hat{y}_1 = \frac{\sum_{u=1}^{n_k} x_i w_i\, g(\left\|\frac{\hat{y}_0-x_i}{h}\right\|^2}{\sum_{u=1}^{n_k} w_i\, g(\left\|\frac{\hat{y}_0-x_i}{h}\right\|^2} \quad (7)$$

Here, $g(x) = -k'(x)$ is taken as the derivative of $k(x)$ for all $x \in [0, \infty)$, except for a finite set of points. The complete object localization algorithm is now presented in the proceeding steps: (1) Present the object model $\{\hat{q}_u\}_{u=1,2,\cdots,m}$ and its location $\hat{y}_0$ in the previous frame, (2) Initialize the location of the object in the current frame with $\hat{y}_0$, compute $\{\hat{p}_u(\hat{y}_0)\}_{u=1,2,\cdots,m}$, and finally evaluate $\rho[\hat{p}(y_0), \hat{q}] = \frac{1}{2}\sum_{u=1}^{m} \sqrt{\hat{p}_u(\hat{y}_0)\hat{q}_u}$. (3) Derive the weights $\{w_i\}_{i=1,2,\cdots,n_k}$, (4) Find the next location of the object candidate, (5) Compute $\hat{p}_u(\hat{y}_1)_{u=1,2,\cdots,m}$ and evaluate $\rho[\hat{p}(y_1), \hat{q}] = \frac{1}{2}\sum_{u=1}^{m} \sqrt{\hat{p}_u(\hat{y}_1)\hat{q}_u}$, (6) While $\rho[\hat{p}(y_1), \hat{q}] < \rho[\hat{p}(y_0), \hat{q}]$ is satisfied, there is $\hat{y}_1 = \frac{1}{2}(\hat{y}_0 + \hat{y}_1)$ toevaluate $\rho[\hat{p}(y_1), \hat{q}]$, (7) If $\|\hat{y}_1 - \hat{y}_0\| < \varepsilon$stop processing, otherwise $\hat{y}_1 = \hat{y}_0$is taken and go to step (3).

2.2.8 Implementation of the Algorithm

Implementation of the proposed tracking algorithm can be much simpler than the presented above. The stopping criterion threshold is initiated, which is derived by constraining the vectors $\hat{y}_0$ and $\hat{y}_1$ to be within the same pixel, in the original image coordinates. A lower threshold may induce sub pixel accuracy. From real-time constraints, i.e., the traditional CPU performance in time, the number of the MS scheme iterations is bounded toN_{max}, typically, taken to be 20. In practice, the average number of iterations is much smaller, about 4. The role of step 5 is only to avoid potential numerical problems in the MS scheme based maximization. These problems can appear due to the linear approximation of the Bhattacharyya coefficient.

However, a large set of experiments in tracking the different objects in lengthy periods of time has shown that the Bhattacharyya coefficients are calculated at the new location $\hat{y}_1$, failed to increase only 0.1% of the cases. Therefore, the Step 5 is not used in practice, and as a result, there is no need to evaluate the Bhattacharyya coefficient in Steps 1 and 4. In the practical algorithm, the weights in Step 2 are iterated, deriving the new location in Step 3, and testing the size of the kernel shift in Step 6. The Bhattacharyya coefficient is resulted only after completing the algorithm to evaluate the similarity between the object model and its chosen candidate. The kernels with Epanechnikov profile is described as

$$k(x) = \begin{cases} \frac{1}{2} C_d^{-1}(d+2)(1-x); & x \leq 1 \\ 0; & \text{Otherwise} \end{cases} \tag{8}$$

In this case, the derivative of the profile $g(x)$ is constant and the result reduces to $\hat{y}_1 = \frac{\sum_{u=1}^{n_k} x_i w_i}{\sum_{u=1}^{n_k} w_i}$, i.e. a simple weighted average. The maximum of the Bhattacharyya coefficient can also be interpreted, as a matched filtering procedure. The MS scheme procedure finds the local maximum of the scalar field of correlation coefficients; due to the use of kernels.

3. The Experiments and Results

The proposed approach performance is now considered through a sequence of different images. At first, the whole of experiments regarding the approach presented here have been entirely implemented via 2013 MATLAB programming language. In one such case, the proposed approach is implemented on a 1.6GHz T2050 under 1GB of RAM, while the frame dimensions are first taken as 466×654 pixels.

In this case, the inputs and the related outputs regarding the present simulation, in its application point of view, are simply related to the number of objects to be chosen and also the number of the same objects to be correspondingly tracked, respectively. Moreover, in the algorithm point of view, the inputs are taken as target model and target candidates, as long as the outputs are taken as the outcomes regarding the similarity function smoothness through Bhattacharyya coefficient and target localization, as well. Figure 2 depicts the hand tracking in the usual image through realizing the MS scheme. Here, four frames including 21, 38, 45 and finally 76 regarding a total of 100 frames have been illustrated. In this case, hand movements are slow and also its resolutions and the corresponding image qualities are somehow acceptable. As is seen, this algorithm could not track the right hand, as an object, properly, in all the simulation time.

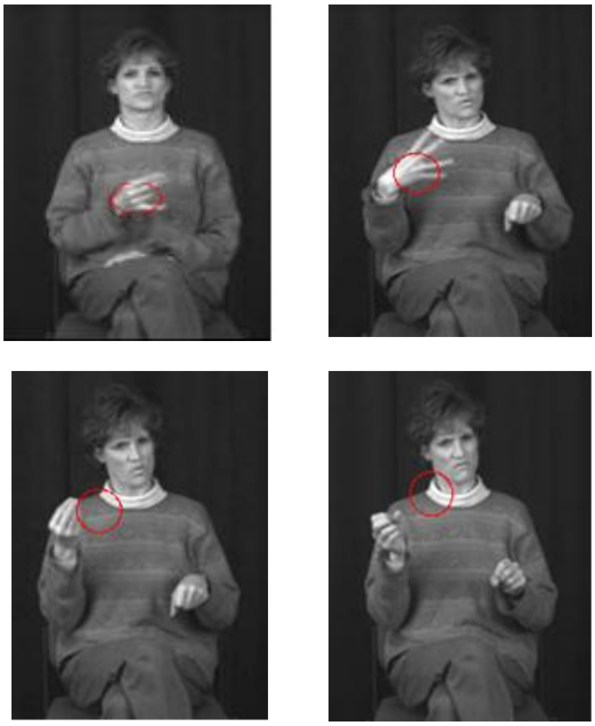

Fig. 2. The hand tracking in a sequence of images though the MS scheme algorithm.

Figure 3 shows the hand tracking in a sequence of usual images through the MD scheme. According to the results, this algorithm could not track just right hand, as an object, accurately. As is obvious, in a number of frames, the face and the left hand have been chosen, as an object. Figure 4 shows the hand tracking, in a sequence of usual images, through the PF scheme tracking the hand but needed lengthy processing time.

Figure 5 illustrates the hand tracking, in the usual image, though the integration of the MD scheme and the MS scheme, as the proposed approach. This one could track just right hand, as an object, in an accurate manner, w. r. t. the mentioned algorithms. Moreover, Fig. 6 illustrates the tracking error in the whole of three considered algorithms. In the MS scheme, object cannot be tracked in all frames because, after several frames, the occlusion happens and its error increases, as well. In the MD scheme, the error does not increase. Not only the right hand is tracked, but also all the moving things are tracked, as is clear to point out.

Fig. 3. The hand tracking in a sequence of images through the MD scheme algorithm.

Finally, in the proposed combination approach; CMP, it can track the hand, so the corresponding errors of all the frames are negligible. Table 1 tabulates the details of the hand tracking in the present above-mentioned algorithms. In all ones, the tracking error has been obtained by subtracting the center of hand and its object.

Fig. 4. The hand tracking in a sequence of images through the PF scheme algorithm

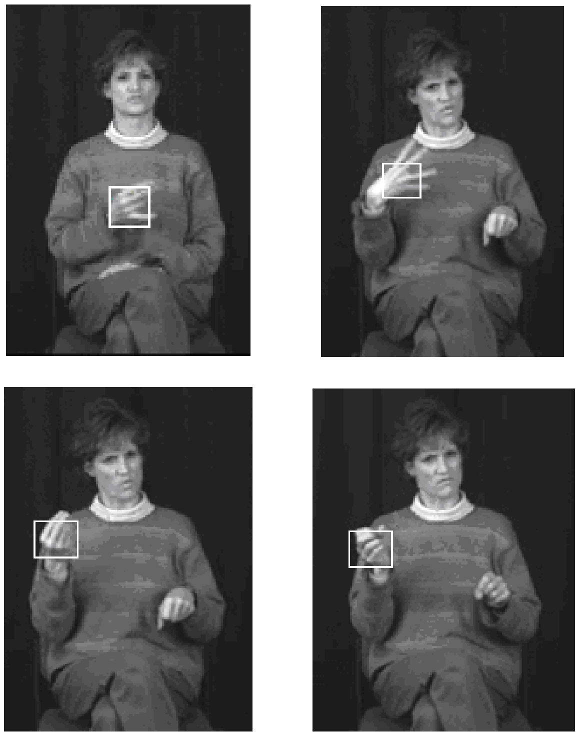

Fig. 5. The hand tracking in a sequence of images through the proposed approach

As is now considered, the processing time of the proposed approach is less than the MD scheme about half. Also, the mean position error in the MD scheme is more than the proposed approach. It is notable that the MS scheme cannot track the hand, continuously, and lose it in the middle of the way.

In one such case, the Bhattacharyya coefficients regarding the experiments, presented in Fig. 5, are illustrated in Fig. 7. It is to evaluate the similarity between the object model and the chosen candidate. As are obvious, this figure indicates that the process of tracking of the chosen object is well behaved. In making an effort to make in considering the effectiveness of the approach proposed here, Fig. 8 is presented in comparison with the MS scheme in the present complicated images sequence. The resolution and its quality regarding the images in this figure are not so desirable and also the processed images are noisy and blurring.

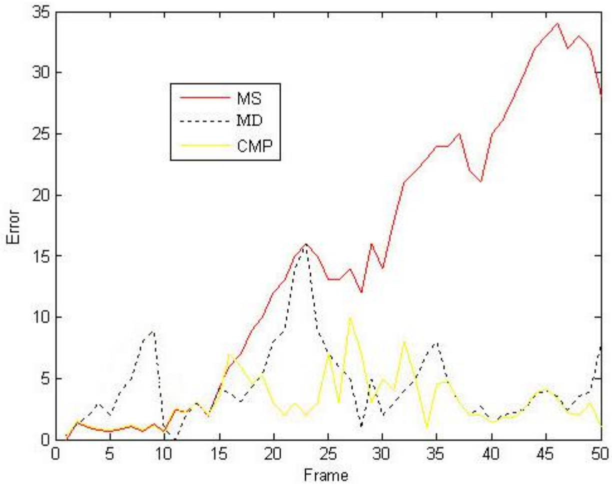

Fig. 6. The tracking error in three algorithms: the MS scheme (red), the MD scheme (dash), the proposed combination (yellow).

And hand movements are faster than the old ones, but, as is seen, this could track the hand, as an object, efficiently. These images have been taken from the forum Iran’s site, although the old pictures are related to the main database, i.e. RWTH-BOSTON-104.

In order to evaluate and compare the proposed approach with a number of conventional methods, a number of tracking techniques including the MD scheme, the MS scheme, the PF scheme and finally the integration of the PF scheme along with the MS scheme are all presented in Table 1 [20]-[22].

Concerning the proposed approach, the integration of the MS scheme along with the MD scheme is realized to be applicable to remove fixed parts. It aims us to be able to track the hand, accurately and efficiently, after choosing the target. Moreover, regarding the specification of the PF scheme, a number of pixels need to be processed, in its lengthy processing time, for each one of incoming frames to realize a tracking system. Hereinafter, regarding the specification of the MS scheme, it is noted that the speed of the present tracking system run time is higher than the mentioned algorithm.

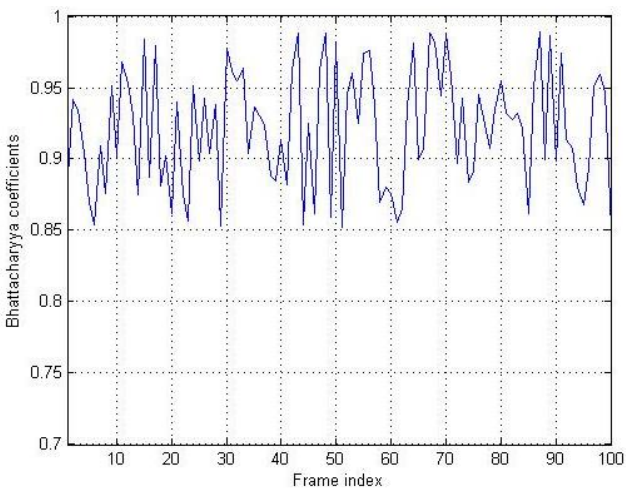

Fig. 7. The Bhattacharyya coefficients in the experiments regarding Fig. 6.

In fact, there are some potential approaches to be considered as comparable methods, while the integration of the PF scheme along with the MS scheme is one of the hybrid solutions, presented.

Now, as considered, both speed of processing time and the tracking accuracy are better than others. Moreover, it should be noted that, the proposed approach has lower complexity, compared to the same conventional tracking methods. Moreover, to consider the effectiveness of the approach proposed w. r. t. the potential benchmarks in this area, at first, four of them are chosen to be analyzed with respect to state-of-the-art. It is apparent that with a focus on the whole of methodologies, investigated in the benchmarks, as well as the potential results, illustrated in [20], the topic of the present research can be explained. It should be noted that the aforementioned reference is a deep collection of the approaches that are directly related to the proposed research topic that is published, in recent year. It aims us to see the new investigations in this field. With regard to the results illustrated there, at first, the idea of the proposed research can be of novelty and therefore its performance is then needed to consider, carefully. It means that the overall performance of the approach proposed here is to be resulted by considering

Fig. 8. The comparison between the MS scheme and the proposed approach in a sequence of the complicated images.

processing time, the accuracy criteria and also the type of environments that were experimentally processed.

Table 1. The comparable outcomes of the proposed approach along with four related tracking schemes.

	Algorithms	Needed pixels	Mean time per frame (msec.)	Mean position error (pixels)
1	The MD scheme	variable	2.5	8
2	The MS scheme	20	1.0	10
3	The PF scheme	150	3.5	8
4	The PF+MS scheme	50	2.5	7
5	The proposed approach	20	1.8	5

Now, the definition of the basic concept for the accuracy criteria (η), and its tracking errors (ℓ), are now given, respectively, by Eq. (9)

$$\begin{cases} \eta = \dfrac{TP_p}{TP_p + FN_p} \\ \ell = \dfrac{\overline{TP_p}}{TP_p + FN_p} \end{cases} \tag{9}$$

Here TP_p is taken as the objects in the frames that have correctly verified, while $\overline{TP_p}$ is taken as the objects in the frames that have mistakenly verified. And finally $TP_p + FN_p$ is the total of entities in the frames that have considered in the process of surveying, as well.

With this purpose, the performance results of the four potential benchmarks along with the proposed approach are now tabulated in Table 2. Now, with regard to the present outcomes, it is wise to note that the superior results with respect to the accuracy outcomes are directly related to the Li's experiments, while the tracking performances of the proposed approach and the Jacob's experiments are located in the second ranking position. The performance of the Kong's experiments and the Kiliboz's experiments are in the third and the forth ranking positions, respectively. Of course, another heuristic factor aims us to finalize the present comparison process. It is also considered to illustrate the complication of the environments that are experimentally processed. It means that the percentage of a very complicated frame is now scored as 100, while other related frames are correspondingly scored below versus this percentage value in order. This is obvious that the more score happens; the more environmental complexity acquires, in the corresponding matter. Now, by considering the whole of factors in one such case, in a careful manner, it is clear to note that the superiority of the proposed approach could somehow be verified in the present cases with respect to other related benchmarks. In making an effort to be realized, in the same way, the processing time regarding the approach presented here in comparison with the whole of mentioned benchmarks are considered. The results in one such case are coherently related to the hardware and the corresponding software performances. Due to the fact that the potential benchmarks are appeared, in recent years, it can be supposed to take the proposed approach hardware specification to be the same as other related ones and therefore the comparisons can be somehow meaningful. In this way, the average processing time consumed per each frame is illustrated in Fig. 9 (in msec.). The present frames processing time is only reported in the Kiliboz's benchmark as 1500 msec. in the about 750 successful processed frames, while this factor in the proposed approach is about 2 msec. The outcomes regarding the time consumed indicate that the frames processing time is somehow competitive w. r. t. the corresponding one.

4. Conclusions

A hybrid MS-MD approach has been now investigated in the present research to organize an insight in the area of hand tracking in high robust performance in the presence of complicated images. The proposed approach uses color features, which can be obtained of the frames, while the MD scheme is realized in partial. The present MD is in fact applied to remove points in association with

the MS scheme to track the hand gesture. The approach proposed here is carried out to track different kind of hands, as the objects, in the real and complicated environments. As is obvious to us, the proposed approach can track the hand gesture that is somehow better than the traditional algorithms, in general. Furthermore, this is in line with some other potential algorithms, realized in the area of object tracking, such as dynamic programming tracking, which is now desirable in real-time domains. Due to the fact that using the MS-MD scheme is somehow efficient and applicable, the less computation is coherently needed to track the object and it is faster than other corresponding algorithms, as it is two times faster than the dynamic programming tracking in the same hardware performance. In addition, less memory is needed to track the object and therefore the outcomes can be useful to implement in some practical environments.

Table 2. The accuracy criteria; η, the tracking error; ℓ, and the frame complication; ρ, considering the benchmarks as well as the proposed approach.

		$\%\eta$	$\%\ell$	$\%\rho$
1	The Kılıboz'sexperiments [15]	73.0	27.0	80
2	The Li's experiments [16]	99.0	1.0	80
3	The Jacob's experiments [22]	92.3	7.7	90
4	The Kong's experiments [23]	89.9	10.1	100
5	The proposed approach	90.4	9.6	90

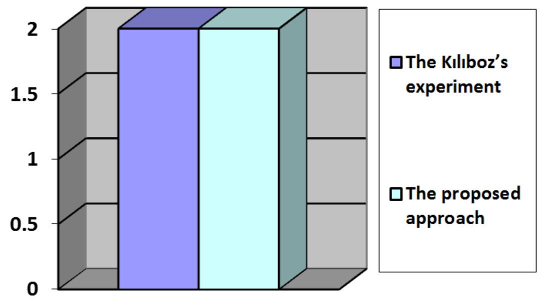

Fig. 9.The average time consumed per each frame (msec.).

Acknowledgements

The corresponding Author would like to express the warmest and best regards to respected Editors of JIST Journal. Moreover, special thanks to all the potential reviewers for the technical, valuable and constructive comments in the process of paper modifications.

References

[1] Tan Dat Nguyen, Surendra Ranganath, Facial expressions in American Sign Language: Tracking and recognition, *Pattern Recognition*, Volume 45, Issue 5, May 2012, Pages 1877-1891.

[2] Qutaishat Munib, Moussa Habeeb, Bayan Takruri, Hiba Abed Al-Malik, American sign language (ASL) recognition based on Hough transform and neural networks, *Expert Systems with Applications*, Volume 32, Issue 1, January 2007, Pages 24-37.

[3] Foti Coleca, Andreea State, Sascha Klement, Erhardt Barth, Thomas Martinetz, Self-organizing maps for hand and full body tracking, *Neurocomputing*, Volume 147, 5 January 2015, Pages 174-184.

[4] Chan-Su Lee, Sung Yong Chun, Shin Won Park, Tracking hand rotation and various grasping gestures from an IR camera using extended cylindrical manifold embedding, *Computer Vision and Image Understanding*, Volume 117, Issue 12, December 2013, Pages 1711-1723.

[5] Malik Morshidi, TardiTjahjadi, Gravity optimised particle filter for hand tracking, *Pattern Recognition*, Volume 47, Issue 1, January 2014, Pages 194-207.

[6] Chien-Ting Huang, Ing-Shiou Hwang, Kinematic property of object motion conditions gaze behavior and eye-hand synergy during manual tracking, *Human Movement Science*, Volume 32, Issue 6, December 2013, Pages 1253-1269.

[7] Yasuyuki Inoue, Yutaka Sakaguchi, Periodic change in phase relationship between object and hand motion during visuo-manual tracking task: Behavioral evidence for intermittent control, *Human Movement Science*, Volume 33, February 2014, Pages 211-226.

[8] Farzad Husain, Babette Dellen, Carme Torras, Robust surface tracking in range image sequences, *Digital Signal Processing*, Volume 35, December 2014, Pages 37-44.

[9] Prashan Premaratne, Sabooh Ajaz, Malin Premaratne, Hand gesture tracking and recognition system using Lucas–Kanade algorithms for control of consumer electronics, *Neurocomputing*, Volume 116, 20 September 2013, Pages 242-249.

[10] S.S. Ge, Y. Yang, T.H. Lee, S.S. Ge, Y. Yang, T.H. Lee, Hand gesture recognition and tracking based on distributed locally linear embedding, *Image and Vision Computing*, Volume 26, Issue 12, 1 December 2008, Pages 1607-1620.

[11] D. González-Ortega, F.J. Díaz-Pernas, M. Martínez-Zarzuela, M. Antón-Rodríguez, J.F. Díez-Higuera, D. Boto-Giralda, Real-time hands, face and facial features detection and tracking: Application to cognitive rehabilitation tests monitoring, *Journal of Network and Computer Applications*, Volume 33, Issue 4, July 2010, Pages 447-466.

[12] Jinshi Cui, Zengqi Sun, Model-based visual hand posture tracking for guiding a dexterous robotic hand, *Optics Communications*, Volume 235, Issues 4–6, 15 May 2004, Pages 311-318.

[13] Sarath Kodagoda, Stephan Sehestedt, Simultaneous people tracking and motion pattern learning, *Expert Systems with Applications*, Volume 41, Issue 16, 15 November 2014, Pages 7272-7280.

[14] Andrea Guazzini, Eiko Yoneki, Giorgio Gronchi, Cognitive dissonance and social influence effects on preference judgments: An eye tracking based system for their automatic assessment, *International Journal of Human-Computer Studies*, Volume 73, January 2015, Pages 12-18.

[15] Nurettin Çağrı Kılıboz, Uğur Güdükbay, A hand gesture recognition technique for human–computer interaction, *Journal of Visual Communication and Image Representation*, in press,2015.

[16] Shao-Zi Li[a, b]Author Vitae,

- Bin Yu[a, b, c]Author Vitae,
- Wei Wu[a, b]Author Vitae,
- Song-Zhi Su[a, b]Author Vitae,
- Rong-Rong Ji[a, b, ,] Author Vitae

[17] Shao-Zi Li, BinYu, WeiWu, Song-ZhiSu, Rong-Rong Ji, Feature learning based on SAE–PCA network for human gesture recognition in RGBD images, *Neurocomputing*, 2015, Pages 565–573.

[18] Siddharth S. Rautaray, Anupam Agrawal, Vision based hand gesture recognition for human computer interaction: a survey, *Artificial Intelligence Review*, Volume 43, Issue 1, January 2015, Pages 1-54.

[19] K. Li, Xu KH, Huang DS (2012) Improved object tracking method based on mean shift and particle filter. J Comput Appl 32(2):504–506.

[20] Qiuyu Zhang, Jingman Liu, Hongxiang Duan, Daodong Wang, Hand gesture tracking based on particle filtering aiming at real-time performance. J Computational Science 10(4):1149–1157.

[21] Kuo-hsien Hsia, S.F. Lien, J. P> Su, Moving target tracking based on camshaft approach and kalman filter, Applied Mathematics Information Sciences, 7(1), 2012, 193-200.

[22] C. F. Shan, T. N. Tan, Y. C. Wei, Real-time hand tracking using a mean shift embedded particle filter, Pattern Recognition, 40(7), 2007, 1958-1970.

[23] Mithun George Jacob, Juan Pablo Wachs, Context-based hand gesture recognition for the operating room, *Pattern Recognition Letters*, Volume 36, 15 January 2014, Pages 196–203.

[24] W.W. Kong, Author VitaeSurendra Ranganath, Author VitaeTowards subject independent continuous sign language recognition: A segment and merge approach, *Pattern Recognition*, Volume 47, Issue 3, March 2014, Pages 1294–1308.

Amir Hooshang Mazinan received the Ph.D. degree in 2009 in Control Engineering. Dr. Mazinan is the Associate Professor and also the Director of Control Engineering Department at Islamic Azad University (IAU), South Tehran Branch, Tehran, Iran. He is now acting as Associate Editor in TIMC Journal (SAGE publisher) and Guest Editor in CAEE Journal (Elsevier Publisher), as well. Moreover, he is a member of Editorial Board in three international journals and also a member of programming committee in four international conferences. Dr. Mazinan has more than 80 journal and conference papers in so many reputable publishers. His current research interests include intelligent systems, model based predictive control, over actuated mechanical system modeling and control, time–frequency representation, filter banks, wavelet theory and image-video processing.

Jalal Hassanian received the M.Sc. degree in 2014 in Electrical Engineering at the Islamic Azad University (IAU), South Tehran Branch, Tehran, Iran. His current research interests include signal, image and video processing.