Video processing for judicial applications

Konstantinos Avgerinakis, Alexia Briassouli, Ioannis Kompatsiaris

Informatics and Telematics Institute,
Centre for Research and Technology, Hellas
Thessaloniki, Greece
{koafgeri,abria,ikom}@iti.gr
http://www.certh.gr/,www.iti.gr

Abstract. The use of multimedia data has expanded into many domains and applications beyond technical usage, such as surveillance, home monitoring, health supervision, judicial applications. This work is concerned with the application of video processing techniques to judicial trials in order to extract useful information from them. The automated processing of the large amounts of digital data generated in court trials can greatly facilitate their browsing and access. Video information can provide clues about the state of mind and emotion of the speakers, information which cannot be derived from the textual transcripts of the trial, and even from the audio recordings. For this reason, we focus on analyzing the motions taking place in the video, and mainly on tracking gestures or head movements. A wide range of methods is examined, in order to find which one is most suitable for judicial applications.

Key words: video analysis, recognition, judicial applications

1 Introduction

The motion analysis of the judicial videos is based on various combinations of video processing algorithms, in order to achieve reliable localization and tracking of significant features in the video. Initially, optical flow is applied to the video, in order to extract the moving pixels and their activity. Numerous algorithms exist for the estimation of optical flow, like the Horn-Schunck developed in 1981 [1] and the Lucas-Kanade [2]. A more recent and sophisticated approach was developed by Bouguet in 2000 [3]. This method implements a sparse iterative version of Lucas-Kanade optical flow in pyramids. By applying the optical flow algorithm, we separate the video frame pixels into a set of points that are moving and a static set of points. As some of the optical flow estimates may be caused by measurement noise, a more accurate separation of the pixels into static and active can be obtained by applying higher-order statistics, namely the kurtosis, to the optical flow data.

After the pixel motion is estimated and the active pixels are separated from the static ones, only the active pixels are processed. This allows the system to operate with fewer errors that would be caused by confusing the noise in static pixels with true motion. The next step in the motion analysis is the interest point

tracking in active pixels. The goal of this stage is to obtain the points which will define the human action. Firstly we want to define the object that appears to move through the frames. This object can be characterized using specific features that will define the image and the object singularly. These specific feature points then need to be matched from frame to frame so as to acquire interest point tracking. Several state of the art algorithms have been examined for the detection and description of these features. These are: The SIFT (Scale Invariant Feature Transform) algorithm which had been published by David Lowe in 1999 [4]. The SURF(Speeded Up Robust Features) algorithm which had been presented by Herbert Bay in 2006 [5], and is based on sums of 2D Haar wavelet responses to make an efficient use of integral images. The Harris-Stephens corner detection algorithm developed in 1988 and finds corners with big eigenvalues in the image [6].

After the feature points are detected, interest point matching is needed. This can be accomplished using a kd-tree algorithm, and specifically the BBF(Best Bin First) algorithm. The BFF finds the closest neighbor of each feature in the next frame based on the two feature's descriptors' distance. In the rest of the paper, the algorithms and their combinations that are used are explained in detail. The results of each technique are shown to demonstrate which ones give the best results.

2 Kurtosis-based Activity Area

In of the methods used in this work, the optical flow estimates can be used immediately, or further processed by higher order statistics in order to obtain a more accurate estimate of the active pixels [7]. The optical flow values may be caused by true motion in each pixel, or by measurement noise, expressed by the following two hypotheses:

$$\begin{aligned} H_0 &: v_k^0(\bar{r}) = z_k(\bar{r}) \\ H_1 &: v_k^1(\bar{r}) = u_k(\bar{r}) + z_k(\bar{r}), \end{aligned} \tag{1}$$

where $v_k^i(\bar{r})$ expresses the flow estimate at frame k and pixel $\bar{r}$ and $i = 0, 1$ depending on the corresponding hypothesis. Also, $u_k(\bar{r})$ is the illumination variation caused by true motion and $z_k(\bar{r})$ is the additive measurement noise. In the literature, additive noise is often modeled by a Gaussian distribution. Thus the separation of the active from the static pixels is reduced to the detection of Gaussianity in the flow estimates. A classical measure of Gaussianity is the kurtosis, which is zero for a Gaussian random variable and is given by:

$$\mathbf{kurt}(\mathbf{y}) = \mathbf{E}[\mathbf{y^4}] - \mathbf{3}(\mathbf{E}[\mathbf{y^2}])^\mathbf{2}. \tag{2}$$

for a random variable y. Thus, once the optical flow is estimated over a sequence of frames, its kurtosis is estimated at each pixel $\bar{r}$, leading to an image of the kurtosis values. The kurtosis image can be easily binarized, as the values of the kurtosis for active pixels are significantly higher than those for static pixels with

noise-induced flow values. This leads to a binary mask showing pixel activity, called the Activity Area, which can be used to isolate the active from the static pixels.

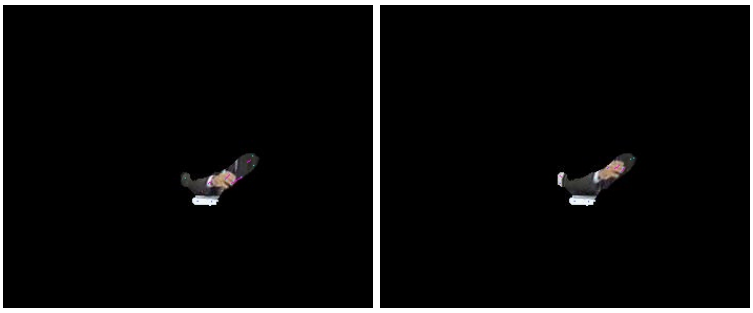
3 Combination of Features and Flow for Active Feature Localization

3.1 HS Optical Flow, Kurtosis, SIFT, BFF Matching

In the first approach examined, the Horn-Schunck algorithm is used to estimate the optical flow of the video [1]. The optical flow values are processed using the kurtosis-based technique of Sec. 2 in order to extract the Activity Area for that video. The Activity Area is shown in Fig. 1(a), and a video frame masked by the Activity Area is shown in Fig. 1(b), from where it can be seen that the moving arm is correctly masked. The Activity Area is used to separate active pixels from the static ones, which are ignored in the rest of the processing. The SIFT algorithm is applied to the active pixels in order to extract features of interest from them [4]. This feature detector is chosen due to its robustness, and in particular because it has been shown to be invariant to image scale and rotation, as well as robust to local occlusion. The resulting features are matched between successive frames, and the results are shown in Fig. 2. In Fig. 2 the blue points indicate which SIFT features have not been matched and the purple lines show the matching between two frames. These results are good but not entirely accurate, as a small number of features is found, and the matching does not provide rich enough information about the activity taking place. Therefore, more methods are examined for increased accuracy in the sequel.

(a) (b)

Fig. 1. (a) Activity Area extracted via the kurtosis method. (b) Video frame masked by the Activity Area.



(a) (b)

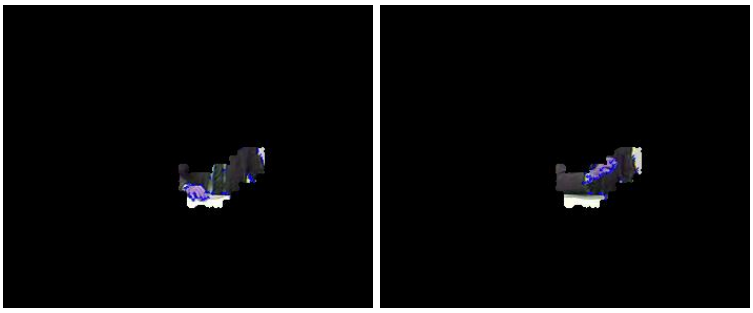
Fig. 2. SIFT points detected on video frames masked by the Activity Area extracted via the kurtosis method.

3.2 Harris Tomas Corner Detection, Lukas-Kanade pyramidal optical Flow, Kurtosis

As features of interest often appear near corners, in this set of experiments the Harris-Thomas Corner Detection is initially used to detect feature points[6]. As Fig. 3 shows, this method provides more feature points than the previously used SIFT feature point detector, and can be therefore considered to give more reliable and robust results. The motion of these feature points also needs to be estimated, to better understand the activity taking place - in this case a gesture. A more recent and sophisticated method for optical flow estimation is used here, namely the pyramidal Lukas-Kanade optical flow, which can deal with both large and small flow values. The kurtosis method is applied again, in order to accurately isolate the truly active features. As Fig. 4 shows, the resulting masked features points provide a good localization of the activity of interest in the video, and can therefore be used in subsequent stages for activity classification or recognition.

(a) (b)

Fig. 3. Harris-Thomas feature points detected on video frames.



(a) (b)

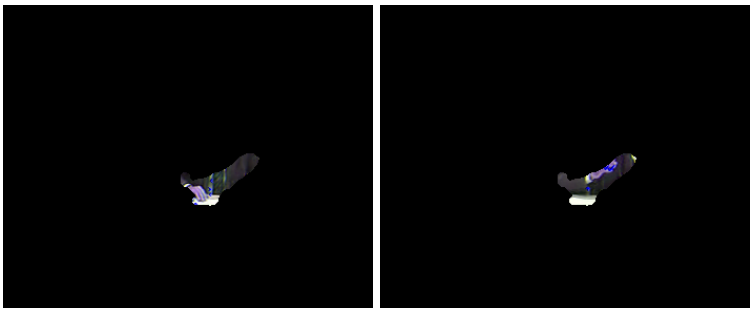
Fig. 4. Harris-Thomas feature points masked by the Activity Area.

3.3 SIFT, Lukas-Kanade pyramidal optical Flow, Kurtosis

In this case, the features are detected using the SIFT algorithm initially, with the results shown in Fig. 5. Afterwards, a more sophisticated method for estimating the motion is used. Namely, the pyramidal Lucas-Kanade is applied to them in order to find their motion throughout the video [3]. The kurtosis of these flow values is estimated so as to eliminate the feature points that are not truly moving, and only keep the active ones. Fig. 6 shows that in this case, fewer features of interest (i.e. moving feature points) are detected than in the method of Sec. 3.2.

(a) (b)

Fig. 5. SIFT feature points.



(a) (b)

Fig. 6. SIFT feature points masked by the Activity Area.

3.4 SURF, Horn-Schunck optical Flow, Kurtosis

In this case, Horn-Schunk optical flow is applied to the pixels where features were detected, and the truly active pixels are extracted, as before, by applying the kurtosis method described above. Afterwards, the resulting Activity Areas are used to mask the video and the resulting, smaller sequence, is input to the SURF algorithm. The SURF [5] method is examined for feature extraction as it has been shown to be more robust than the SIFT algorithm against a wider range of image transformations. Additionally, it runs significantly faster, which is important when processing many long video streams. Features on the active pixels are then found by SURF and, as Fig. 7 shows, they can successfully isolate the most interesting active parts of the video.

(a) (b)

Fig. 7. SURF feature points found in the video after it is masked by the Activity Area.

4 Conclusions

A variety of video processing algorithms has been applied to judicial videos containing characteristic gestures, in order to determine which ones are most appropriate for isolating the activity of interest. The optical flow is extracted as the moving points are of interest in this work. The active pixels are separated from the static ones using a kurtosis-based method. The SIFT algorithm is initially tested for feature as it is known to be robust, but is shown to provide too sparse noiseless feature points for an accurate depiction of the activity taking place. The Harris-Thomas detector and the SIFT algorithms, combined with a pyramidal version of the Lukas-Kanade algorithm, provide an accurate representation of the features of interest in the video. Finally, the SURF algorithm is examined, as it is one of the current state of the art methods for feature extraction, being robust to a wide range of transformations and computationally efficient. The results of applying SURF to the active pixel areas are very satisfactory and can be considered as reliable information for activity classification at latter stages. Future work includes examining additional feature detectors like the STAR detector, as well as performing feature point matching for the detectors examined.

Acknowledgements

The research leading to these results has received funding from the European Community's Seventh Framework Programme FP7/2007-2013 under grant agreement FP7-214306 - JUMAS.

References

1. Horn B.K.P., Schunck B. G.: Determining Optical Flow. Artificial Intelligence 17, 185-203 (1981)
2. Lucas B., T. Kanade: An Iterative Image Registration Technique with an Application to Stereo Vision. In: Proc. of 7th International Joint Conference on Artificial Intelligence (IJCAI), pp. 674-679. (1981)
3. Bouguet J.: Pyramidal Implementation of the Lucas Kanade Feature Tracker. In: OpenCV distribution. (2000)
4. Lowe D. G.: Distinctive Image Features from Scale-Invariant Keypoints. International Journal of Computer Vision. 60, 91–110 (2004)
5. Smith, Bay H., Ess A., Tuytelaars T., Van Gool L.: SURF: Speeded Up Robust Features. Computer Vision and Image Understanding. 110, 346–359 (2008)
6. Harris C., Stephens M.: A combined corner and edge detector. In: Proceedings of the 4th Alvey Vision Conference, pp. 147–151. (1988)
7. Briassouli A., Kompatsiaris I.: Robust Temporal Activity Templates Using Higher Order Statistics. IEEE Transactions on Image Processing. To appear.